

# Se lancer avec l'IA en entreprise : guide pour les débutants



**Red Hat**

# Sommaire



## Introduction

page 3

Chapitre 1

## L'essor de l'IA

page 6

Chapitre 2

## Le choix du modèle d'IA

page 9

Chapitre 3

## Prérequis pour se lancer

page 12

Chapitre 4

## Adoption et mise à l'échelle avec Red Hat

page 16

En savoir plus

## Envie de passer à l'étape suivante de votre parcours d'adoption ?

page 20

## Introduction

### Les entreprises ont de plus en plus conscience des opportunités offertes par l'intelligence artificielle (IA) dans tous les aspects de leur activité.

De l'engagement client à l'assistance et aux ventes, en passant par l'infrastructure informatique, les processus, le développement de code et la distribution de solutions, l'IA est présente dans un nombre croissant de cas d'utilisation et connaît un essor dans tous les secteurs.

D'après IDC, ce marché devrait dépasser les 423 milliards de dollars d'ici 2027 avec un taux de croissance annuel composé de 26,9 % sur 5 ans, un résultat dû aux nombreuses entreprises qui cherchent à améliorer l'efficacité de l'exploitation, l'expérience client et la productivité à l'aide de l'IA<sup>1</sup>.

L'évolution rapide de cette technologie exerce une pression considérable sur les dirigeants qui doivent se dépêcher d'identifier, de choisir, de concevoir et de distribuer des solutions d'IA qui procureront un avantage concurrentiel à leur entreprise. Cependant, toutes les entreprises ne développent pas l'IA au même rythme, et la fréquence des innovations varie également. Dans un tel contexte, difficile de tirer pleinement parti de cette technologie qui soulève souvent plus de questions qu'elle n'apporte de réponses.



Que vous entamiez tout juste votre adoption de l'IA, que vous souhaitiez comprendre son impact potentiel sur votre entreprise ou que vous cherchiez à mettre à l'échelle vos mises en œuvre actuelles, vous trouverez dans ce livre blanc la réponse à de nombreuses questions qui entourent l'IA aujourd'hui.

<sup>1</sup> IDC FutureScape Webcast, « [Worldwide Artificial Intelligence and Automation 2024 Predictions](#) », document n° US51901124, mars 2024

## Les différents types d'IA

Pour tirer le meilleur parti de l'IA, vous devez tout connaître à son sujet, notamment deux des principaux types utilisés par les entreprises à l'heure actuelle.

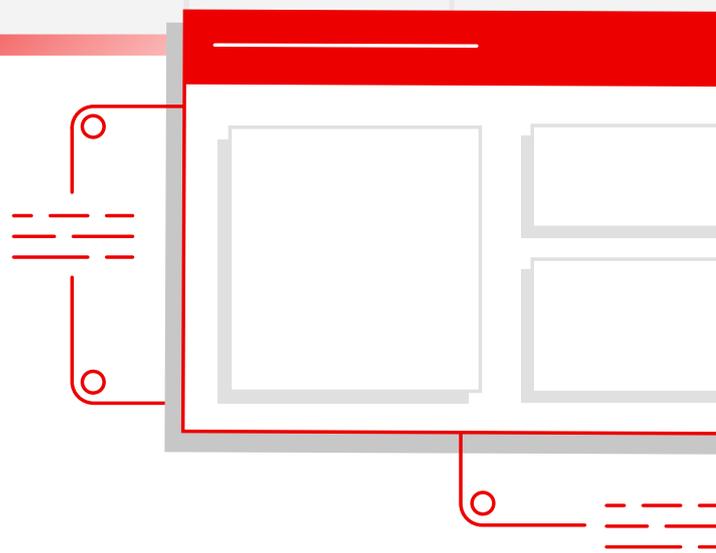
**IA prédictive :** à l'aide de données d'historique, l'IA prédictive aide les entreprises à identifier des modèles et à prendre des décisions éclairées concernant l'avenir. Les modèles prédictifs sont au cœur d'applications telles que les prévisions de la demande, la maintenance prédictive et la planification de l'exploitation. L'IA prédictive s'appuie sur des techniques éprouvées de science des données et d'apprentissage automatique (AA), ce qui lui permet de s'améliorer à mesure que le volume de données traitées augmente.



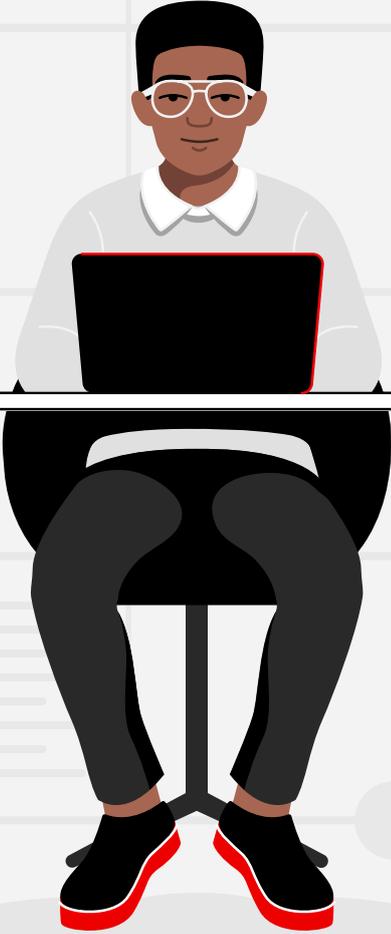
**IA générative :** alimentée par des modèles d'apprentissage profond tels que les transformeurs, l'**IA générative** est capable de créer du contenu, par exemple du texte, des images et du code. Elle s'avère particulièrement utile pour les applications telles que les chatbots, la génération automatisée de contenu et les outils de création. Les modèles de type GPT (Generative Pre-Trained Transformer) ont révolutionné le traitement du langage naturel et la création de contenu en produisant du texte et des images comparables à ce qu'un humain peut produire.

## Avantages de l'IA

Si l'étendue des capacités de l'IA reste à découvrir, il est intéressant de comprendre comment cette technologie en pleine évolution profite d'ores et déjà à d'autres entreprises de toutes tailles et de tous secteurs pour savoir où l'intégrer.



Voici quelques bénéfices de l'IA, ainsi que des pistes pour en tirer parti :



1010  
11011

**Volume de données.** En raison de la croissance exponentielle des données qu'elles collectent, les entreprises peinent souvent à les gérer et à en extraire des informations exploitables. L'IA peut traiter et analyser rapidement de gros volumes de données pour en tirer des informations et des tendances précieuses qu'il serait difficile d'identifier manuellement.

📊 ⚙️

**Efficacité de l'exploitation.** De nombreuses entreprises savent que les processus inefficaces et les goulets d'étranglement peuvent nuire à la productivité et nécessiter davantage de temps et d'efforts. L'automatisation basée sur l'IA rationalise l'exploitation. Résultats : moins d'erreurs et des processus plus efficaces. Elle sert notamment à générer automatiquement des notes de réunion avec une liste des mesures à prendre et des étapes claires, ou à accélérer la création de visuels et de contenus vidéos pour les sites web ou les réseaux sociaux.

👤

**Attentes des clients.** Les clients recherchent des expériences simples et personnalisées. En analysant les données client et en proposant des recommandations et des interactions personnalisées, l'IA contribue à optimiser le service clientèle et à personnaliser les services.

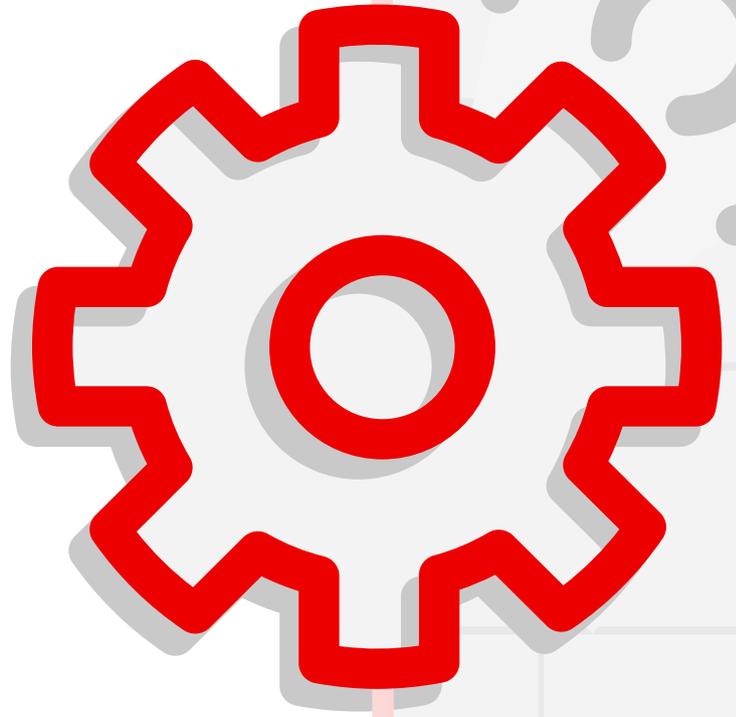
💡

**Compétitivité sur le marché.** Pour rester compétitives au sein d'un marché en rapide évolution, les entreprises ont besoin d'innover en continu. L'IA peut les aider à s'adapter sans délai aux changements du marché et à conserver leur avantage concurrentiel. Les dirigeants peuvent également affiner leur approche en utilisant l'IA générative comme outil de réflexion ou pour préparer une réunion importante.

# L'essor de l'IA

**L'IA évolue depuis des décennies et permet des avancées majeures dans des secteurs tels que la santé, la finance et la fabrication.**

L'essor récent de l'IA générative a attiré l'attention en raison de sa capacité à créer des textes, des images et même du code logiciel dignes d'un humain. Contrairement à l'IA traditionnelle qui automatise des tâches ou analyse des données, l'IA générative offre de nouvelles perspectives en matière de résolution créative des problèmes et de création avancée de contenus.



## Les types de modèles d'IA qui accélèrent l'innovation

Les **grands modèles de langage (LLM)** et les modèles de diffusion latente (ou « stable diffusion ») font partie des modèles d'IA responsables de la croissance fulgurante de l'IA générative. Les LLM, comme GPT, sont préentraînés à l'aide de jeux de données massifs. Ils peuvent comprendre le langage naturel et en générer, ce qui représente un atout précieux pour l'automatisation de l'assistance clientèle, la génération de textes marketing et plus encore. De leur côté, les modèles de diffusion latente créent des images ultraréalistes et stimulent l'innovation dans les domaines du divertissement, du marketing, etc.

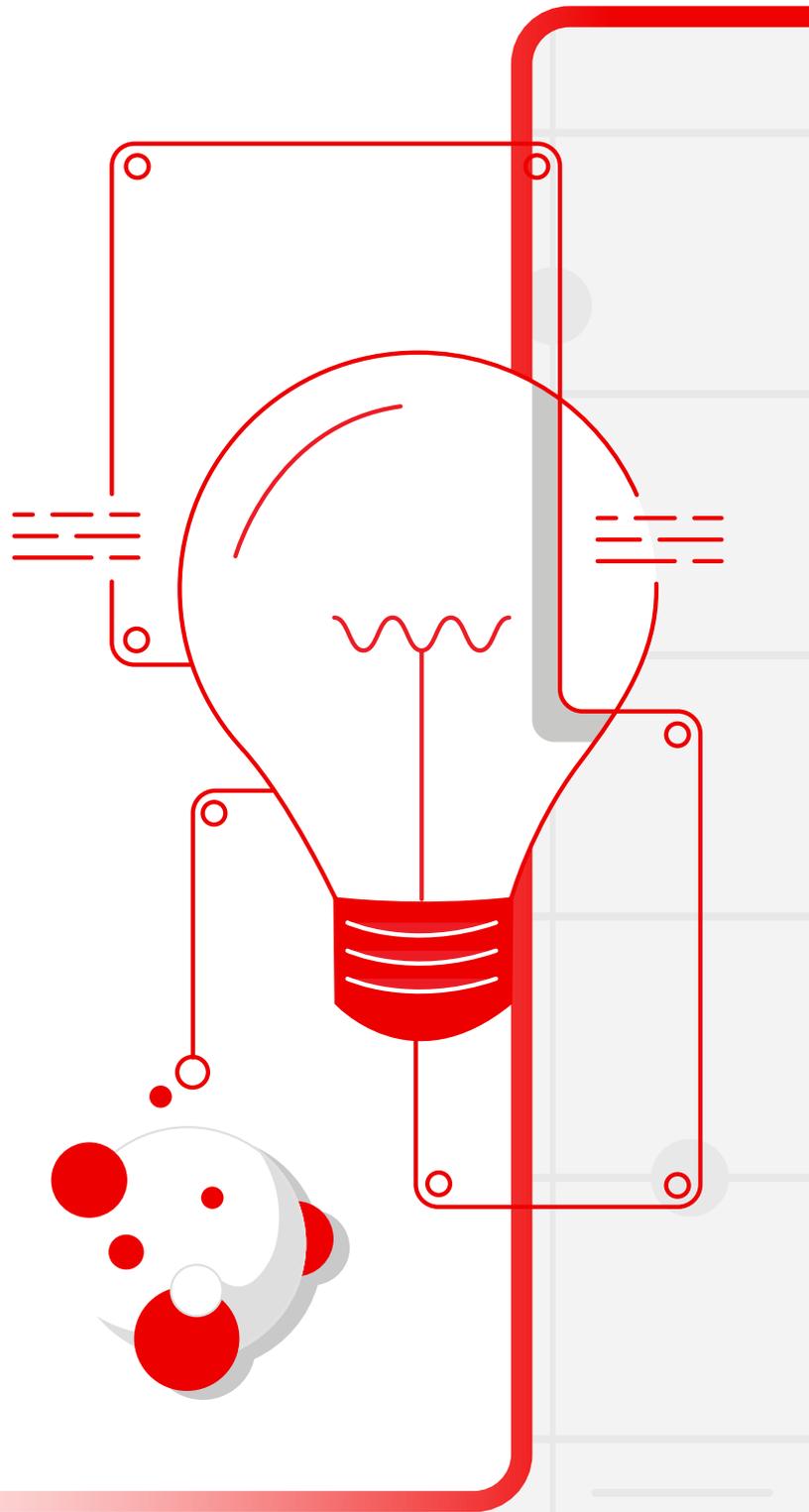
## Les tendances émergentes

Les entreprises explorent de plus en plus l'IA multimodale, qui associe des capacités de traitement de texte, des images et des données au sein d'un même modèle afin d'offrir des solutions plus polyvalentes. Il s'agit d'une tendance à ne pas négliger pour exploiter au maximum le potentiel de l'IA dans le contexte de l'entreprise.

## L'Open Source comme base de l'innovation avec l'IA

**Notre stratégie en matière d'IA** est profondément ancrée dans l'Open Source et permet aux entreprises d'adopter l'IA générative avec plus de transparence et de confiance tout en réduisant les coûts. À l'aide de nos plateformes de **cloud hybride** ouvert, les entreprises peuvent innover librement et garder le contrôle de leurs solutions d'IA.

[En savoir plus sur les LLM et leur fonctionnement](#)



## Prenez le contrôle des LLM grâce à l'Open Source

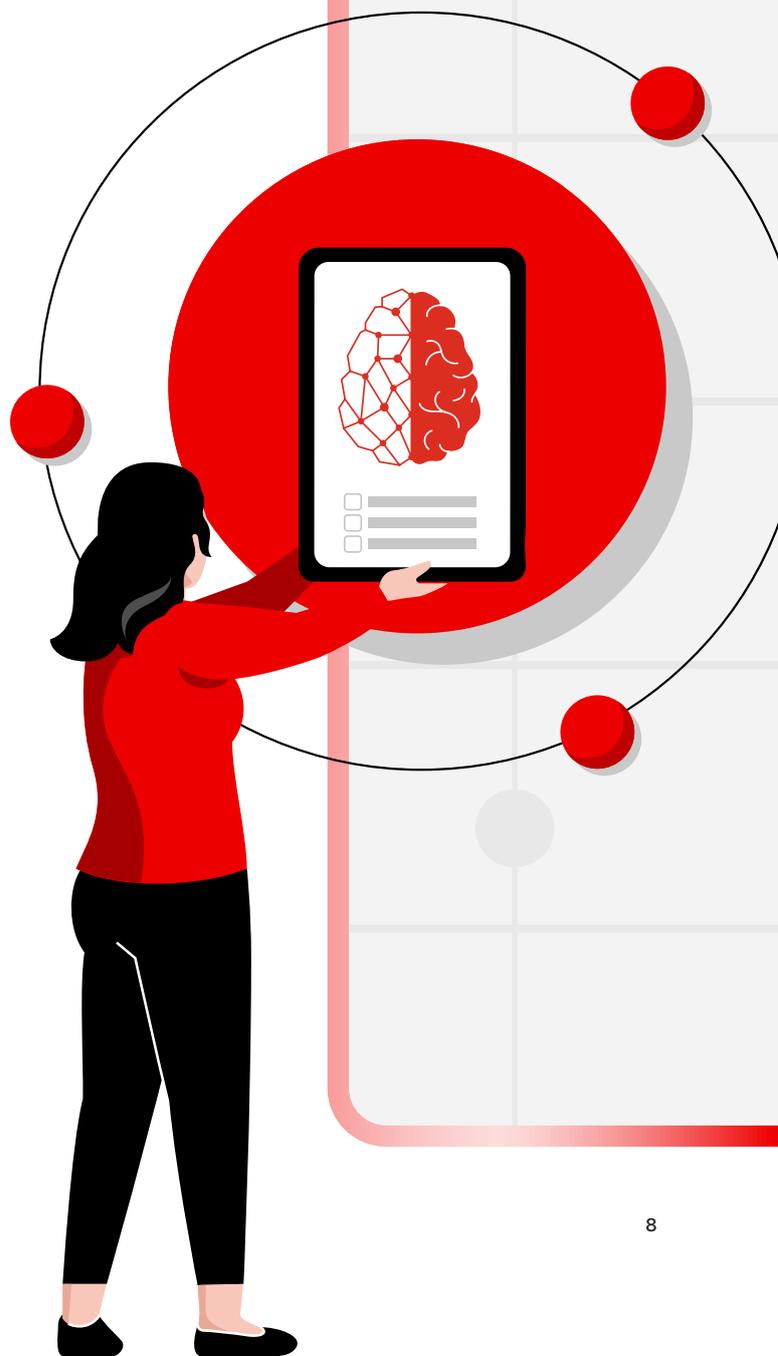
Si l'IA générative refaçonne quasiment tous les aspects de l'activité économique, de la conception des logiciels à notre manière de communiquer, il n'est pas rare que les modèles (LLM ou autres) utilisés dans le cadre de l'IA générative soient étroitement contrôlés par le fournisseur du service. Ainsi, il est difficile pour une entreprise d'évaluer les capacités d'un service d'IA générative sans connaissances spécialisées et sans y consacrer à la fois du temps et de l'argent.

Le manque de visibilité sur les jeux de données qui alimentent un modèle ou sur la manière dont ces données sont utilisées expose les entreprises à des risques liés au contenu généré par IA. Que se passe-t-il si un modèle de génération de code a été entraîné à partir de code source propriétaire ? Le code généré par ce modèle est-il aussi propriétaire ? Ces questions restent ouvertes. Toutefois, les entreprises qui ont compris l'importance potentielle des conséquences se tournent petit à petit vers l'IA Open Source.

Notre approche en matière d'IA est ancrée dans l'Open Source, ce qui explique notre soutien à différents modèles Open Source tels que la famille de modèles de fondation [IBM Granite](#).

Par ailleurs, nos solutions d'IA contribuent directement au développement des modèles d'IA par le biais d'[InstructLab](#), une solution communautaire visant à améliorer les capacités des LLM.

[Découvrir InstructLab sur GitHub](#)



# Le choix du modèle d'IA

## Il existe différents modèles d'IA conçus pour différents cas d'utilisation.

Il est possible d'utiliser à la fois des modèles d'IA prédictive et d'IA générative au sein d'une même application ou d'un même service. Chacun de ces modèles entraînera différents coûts et offrira des avantages uniques, mais peut aider les entreprises à réduire le délai de mise sur le marché pour la preuve de concept initiale. Parmi les exemples les plus utilisés et les plus performants figurent les modèles de segmentation d'images, de conversion de la voix en texte et de reconnaissance d'images, mais le plus important reste d'identifier la meilleure solution en fonction de votre cas d'utilisation.

Entraînés à partir de gros volumes de données, les modèles de fondation offrent des capacités très flexibles. Toutefois, leur taille importante risque d'accroître les coûts, les efforts de gestion et la complexité, ce qui signifie qu'ils ne conviennent pas nécessairement à tous les cas d'utilisation.

Les modèles plus petits et à réglage fin (toujours dans le cadre de l'IA générative) peuvent constituer une meilleure solution pour les entreprises qui cherchent plutôt à affiner leur modèle existant. Il existe également des modèles prêts à l'emploi, faciles à intégrer aux systèmes déjà en place, tels que les LLM. Il s'agit d'outils puissants déjà entraînés à partir de grands ensembles de données.

Néanmoins, si vous devez répondre à des exigences métier spécifiques ou résoudre des problèmes de confidentialité des données, ou si vous souhaitez contrôler davantage le comportement du modèle, il sera peut-être plus judicieux de créer et d'héberger vous-même un modèle personnalisé.

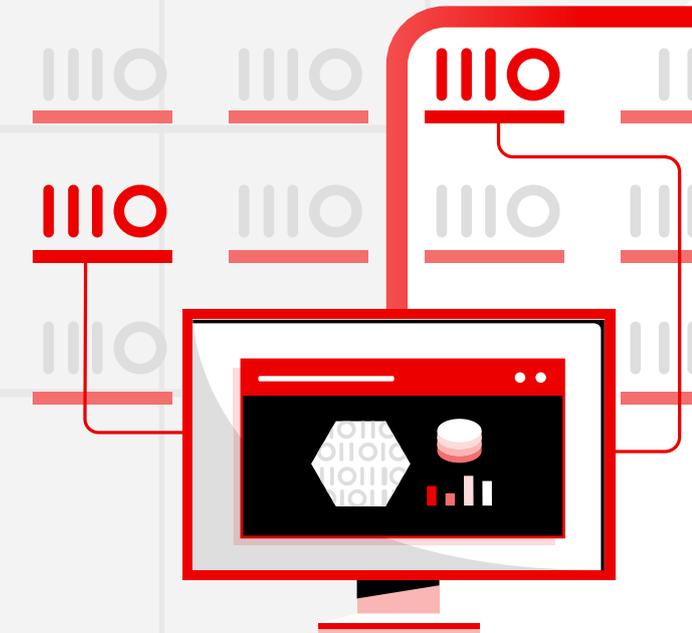
## Comparaison entre la création et le réglage d'un modèle

La création d'un modèle d'IA de A à Z peut représenter un défi de taille. D'abord, vous devez rassembler et préparer de grands jeux de données qui répondent aux cas d'utilisation métier de votre entreprise. Ensuite, il faut choisir un algorithme adapté et l'entraîner avec vos données. Ce processus nécessite une puissance de calcul considérable ainsi qu'une solide expertise, ce qui le rend à la fois chronophage et gourmand en ressources. Si la création d'un modèle de fondation traditionnel permet d'obtenir une solution personnalisée, il ne s'agit pas toujours de l'option la plus efficace.

Le réglage d'un modèle de fondation implique d'adapter un modèle préentraîné à vos exigences spécifiques. L'apprentissage par transfert est une approche courante qui consiste à prendre un modèle entraîné à partir d'un vaste jeu de données pour le réentraîner à l'aide d'un ensemble de données plus petit et propre à un domaine. Avec cette méthode, le modèle conserve les connaissances générales acquises lors de son entraînement initial tout en s'adaptant aux nuances des données qui vous intéressent.

## Réglage fin d'un modèle

Le réglage fin est une autre approche qui consiste à ajuster les paramètres, ou variables, d'un modèle pour améliorer ses performances sur une tâche précise. C'est en intégrant ses données internes au modèle que l'entreprise peut identifier les paramètres à modifier pour obtenir le résultat recherché. Le réglage fin peut consister à changer le taux d'apprentissage ou l'architecture du modèle, ou à entraîner certaines couches de manière plus intensive que d'autres. Ces techniques contribuent à améliorer les connaissances du modèle pour qu'il soit mieux adapté à votre cas d'utilisation précis.



**InstructLab** mise sur une approche de réglage fin, en vue de réduire les connaissances préalables en matière d'IA et de simplifier l'ajout de connaissances d'entreprise aux modèles de fondation Granite existants.

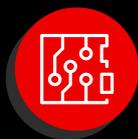
## Autres solutions pour le réglage de modèles

Des chercheurs tentent de trouver une façon de régler les modèles de fondation afin d'accroître leur vitesse et leur efficacité. Parmi les techniques couramment utilisées figure la **génération augmentée de récupération (RAG)**, qui permet de récupérer des faits à partir d'une source externe dans laquelle des connaissances supplémentaires, c'est-à-dire du contexte, ont été ajoutées.

La RAG s'appuie sur l'utilisation d'une ou de plusieurs bases de données externes (bases de données vectorielles), qui apportent du contexte aux questions posées au modèle d'IA générative. Autre approche émergente : l'IA agentique. Elle associe plusieurs agents d'IA générative pour interroger des systèmes externes, tels que des bases de données internes, des intranets d'entreprise ou Internet, afin de fournir les informations les plus précises et actuelles aux modèles d'IA générative.

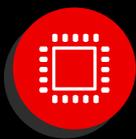
Citons pour finir la technique du **« prompt tuning »**. Il s'agit de fournir aux modèles d'IA du contexte ou des prompts front-end contenant des mots supplémentaires ou des nombres générés par une IA, afin de les orienter vers le résultat souhaité. En alliant « prompt tuning » et RAG, il est possible d'obtenir du contexte supplémentaire pour un nouveau prompt à partir de la réponse donnée à une question RAG. L'association du réglage fin sur un LLM existant aux techniques de RAG et au « prompt tuning » permet aux entreprises qui disposent d'un volume limité de données de personnaliser un modèle de fondation en fonction d'une tâche très précise.

L'infrastructure à la base du modèle d'IA est aussi importante que le modèle lui-même. Chaque tâche nécessite un type de matériel différent.



#### Processeur

Processeurs traditionnels qui gèrent les tâches de calcul générales. Bien que polyvalents, ils ne s'avèrent pas forcément efficaces pour les charges de travail d'IA à grande échelle.



#### Processeur graphique

Processeurs spécialisés, conçus pour gérer des tâches de traitement parallèles. Les processeurs graphiques (GPU) conviennent parfaitement à l'entraînement de modèles d'apprentissage profond qui impliquent le traitement simultané de vastes quantités de données.



#### Processeur neuronal

Nouveau type de processeur spécialement conçu pour les tâches d'IA et offrant davantage d'efficacité et de rapidité pour certains types de modèles.

## Le rôle du cloud hybride dans l'adoption de l'IA par les entreprises

Les environnements de cloud hybride jouent un rôle essentiel dans l'adoption de l'IA. Un cloud hybride associe une infrastructure sur site à des ressources dans un cloud public ou privé, ce qui permet de déployer et gérer les charges de travail d'IA où et comme vous voulez. Par exemple, vous pouvez entraîner vos modèles d'IA à l'aide de GPU puissants basés dans le cloud, puis les déployer sur site ou dans un cloud privé pour des raisons de sécurité ou de conformité. Par conséquent, il est essentiel de veiller à la cohérence des outils et de la plateforme de cloud hybride que vous choisissez.

Notre approche ouverte en matière de cloud hybride aide les entreprises à intégrer l'IA dans divers environnements pour améliorer la cohérence, l'évolutivité et la flexibilité. Cette approche permet de gérer les charges de travail d'IA dans plusieurs environnements cloud, sur site ou en périphérie du réseau, d'optimiser le positionnement des données et de simplifier la migration des données, ce qui facilite l'adoption de l'IA à l'échelle de l'entreprise.

Grâce à une bonne compréhension des modèles d'IA, des données et des infrastructures, vous pouvez surmonter plus facilement les difficultés liées à l'adoption de l'IA et exploiter tout le potentiel de cette technologie.

# Prérequis pour se lancer

L'adoption de l'IA, comme de toutes autres technologies, présente certaines difficultés.

Suivez les étapes ci-dessous pour évaluer l'état de préparation de votre entreprise et identifier les domaines sur lesquels axer vos efforts afin d'accélérer l'adoption de l'IA.



**Évaluation de la qualité et de la disponibilité des données.** L'efficacité d'une IA dépend avant tout de la pertinence et de la qualité des données à disposition. Vérifiez l'exhaustivité, l'exactitude et l'adéquation de vos données afin d'optimiser l'entraînement de vos modèles.

**Évaluation de l'infrastructure technologique.** Vérifiez que votre infrastructure actuelle est en mesure de gérer les charges de travail d'IA. Évaluez notamment la disponibilité des ressources de calcul hautes performances, les solutions de stockage, l'automatisation et les capacités réseau.

**Identification des compétences requises.** Identifiez les compétences en IA déjà présentes dans l'entreprise ainsi que les lacunes qui pourraient nécessiter une formation ou l'embauche d'un spécialiste.

**Analyse de l'alignement stratégique.** Assurez-vous que vos initiatives en matière d'IA sont en phase avec vos objectifs et vos stratégies métier. Les projets d'IA doivent soutenir les objectifs stratégiques globaux de votre entreprise et générer de la valeur métier mesurable.

## Se lancer avec l'IA

De nombreux facteurs influent sur la rapidité et la capacité des entreprises à adopter l'IA. Toutefois, quel que soit le projet de modernisation technologique, il est souvent plus intéressant de commencer à petite échelle et d'avancer progressivement.

Voici huit étapes pour vous lancer et progresser dans votre parcours d'adoption de l'IA :

1

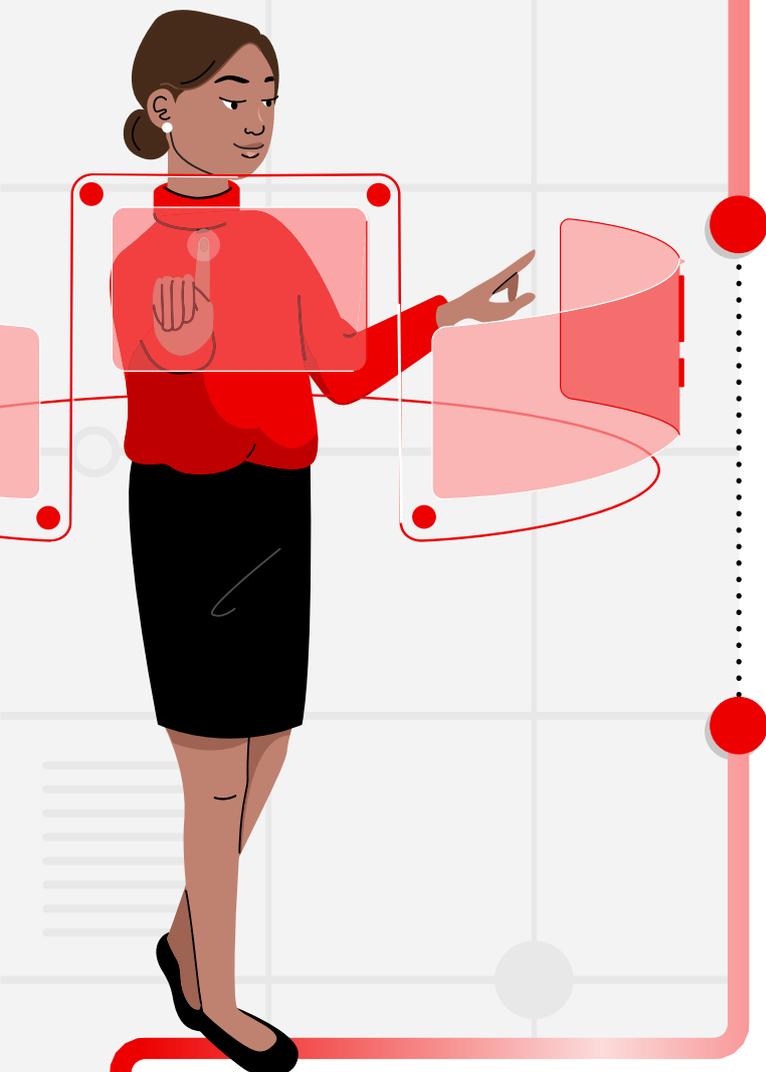
### Évaluation des capacités et des objectifs

Commencez par évaluer les capacités, l'infrastructure et les objectifs stratégiques actuels de votre entreprise. Déterminez si l'IA est en phase avec vos objectifs globaux et identifiez les domaines dans lesquels elle peut apporter de la valeur. Cette étape initiale permettra de définir un cap clair pour votre parcours d'adoption.

2

### Identification des cas d'utilisation et création des équipes chargées de l'IA

Identifiez les opportunités d'utilisation de l'IA au sein de votre entreprise. Formez une équipe pluridisciplinaire composée de développeurs, de spécialistes de l'IA, de data scientists et de spécialistes en informatique pour porter vos initiatives en matière d'IA. Avec un cas d'utilisation bien défini, il est plus facile d'orienter correctement vos efforts d'adoption et de guider vos ressources.



# 3

## Sélection du modèle

Choisissez le modèle d'IA qui convient le mieux à votre cas d'utilisation. Qu'il s'agisse d'un LLM pour l'IA générative ou d'un modèle prédictif pour l'analyse de données, veillez à ce que ses capacités soient en phase avec vos objectifs. Tenez compte des facteurs tels que la complexité et l'évolutivité du modèle, ainsi que sa compatibilité avec vos systèmes existants.

# 4

## Boucles de tests et de validation

Définissez des critères de réussite clairs pour la mise en œuvre de l'IA, notamment des indicateurs de mesure des performances, des taux de précision ou des objectifs métier. Mettez en place des boucles de tests et de validation pour évaluer l'efficacité de votre modèle en continu. Les retours réguliers issus de ces boucles vous aideront à régler finement le modèle et à progresser dans votre parcours d'adoption.

# 5

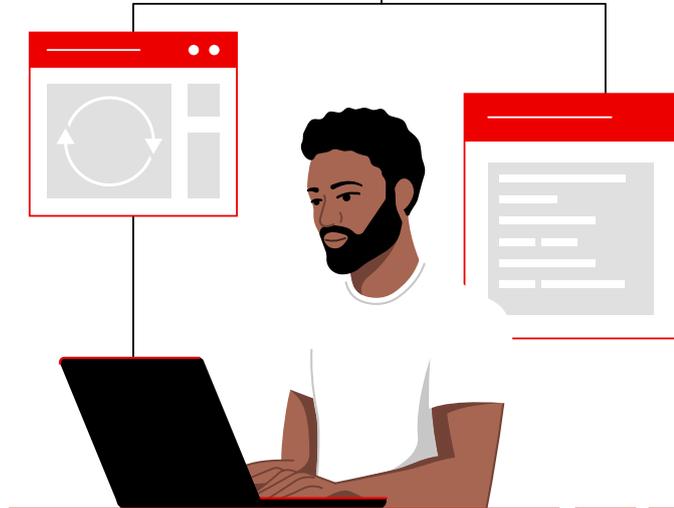
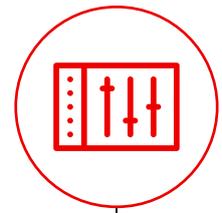
## Réglage du modèle

Personnalisez votre modèle à l'aide des données de votre entreprise. Ce processus implique d'alimenter le modèle avec des données pertinentes pour améliorer sa précision et son adéquation avec votre cas d'utilisation métier. Le réglage fin permet d'adapter le modèle au contexte et aux besoins uniques de votre entreprise.

# 6

## Entraînement à partir de données synthétiques

Envisagez l'utilisation de données synthétiques pour entraîner et améliorer davantage vos modèles d'IA. Cette approche s'appuie sur des processus tels que la distillation des modèles et vous permet de générer des données d'entraînement de haute qualité lorsque les données réelles sont sensibles ou insuffisantes. Ces données synthétiques contribuent à améliorer la robustesse et les performances du modèle sans compromettre la confidentialité.



# 7

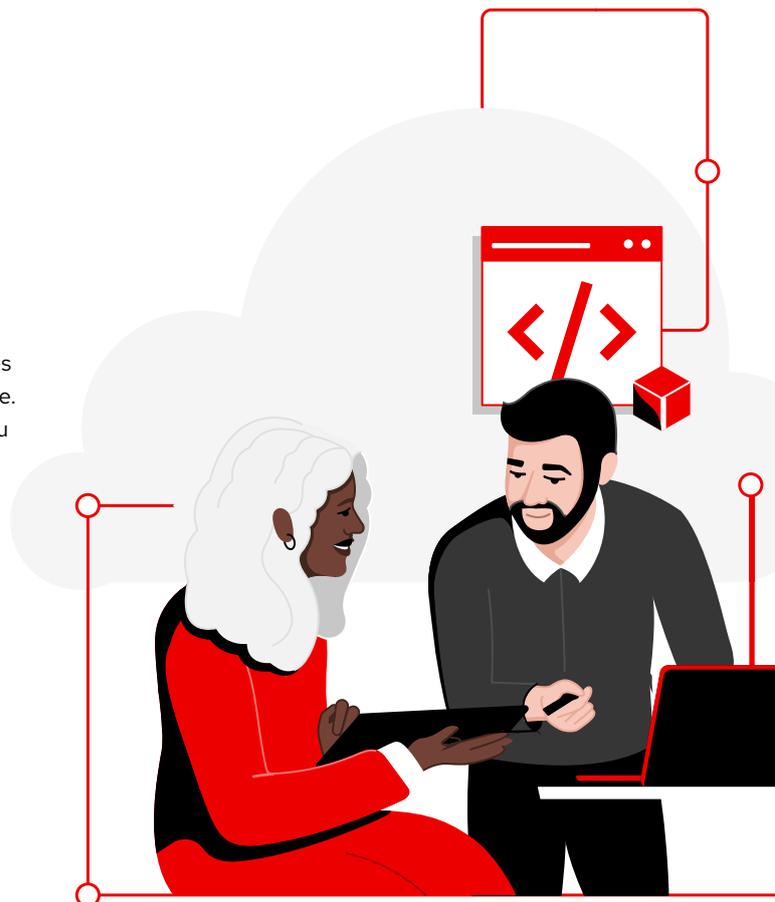
## Surveillance de la dérive

Ce processus permet de réaliser une surveillance générale et basée sur le contenu. Plutôt que de gérer la configuration de manière structurée, la surveillance de la dérive permet de suivre les modifications du contenu au niveau du système de fichiers local. Elle peut vous aider à détecter et traiter tous les écarts du modèle ainsi que les baisses de qualité à long terme. La surveillance continue assure la pertinence et l'efficacité du modèle malgré l'évolution des conditions.

# 8

## Recours à l'aide de spécialistes

Si vos compétences internes en matière d'IA ne sont pas suffisantes, pensez à solliciter l'aide de spécialistes externes, notamment de l'équipe de consulting Red Hat. Ces experts peuvent vous guider tout au long du processus d'adoption de l'IA, fournir des informations précieuses et proposer des formations. Ils peuvent également accélérer votre parcours d'adoption et augmenter vos chances de réussite.



## L'adoption de l'IA repose sur la collaboration

Pour mener à bien vos projets d'IA générative, il est nécessaire de créer une équipe pluridisciplinaire<sup>2</sup>

- Les **responsables métier** utilisent la solution ou profitent de ses retombées.
- Les **spécialistes de l'IA** ajustent les modèles d'IA générative et en assurent la maintenance ainsi que la mise à jour.
- Les **data scientists** traitent les données d'entraînement des modèles en amont pour garantir leur précision et éviter les biais.
- Les **responsables de l'éthique et de la conformité** veillent à ce que tous les projets d'IA générative respectent les réglementations applicables.
- Les **spécialistes de l'exploitation** intègrent les solutions aux infrastructures existantes et mettent en œuvre des politiques de sécurité.
- Les **équipes et communautés de développement** doivent être impliquées au début du projet afin de pouvoir créer, partager et améliorer les outils Open Source, les frameworks et les meilleures pratiques d'adoption de l'IA. Ainsi, les cas d'utilisation de l'IA seront axés sur la génération de valeur métier.

<sup>2</sup> Kearney, « [Standing up tiger teams to tackle generative AI complexity](#) », 15 novembre 2023

# Adoption et mise à l'échelle avec Red Hat

**Red Hat AI assure la fiabilité et la cohérence du cloud hybride et propose suffisamment de choix pour accélérer l'adoption de l'IA dans l'entreprise.**

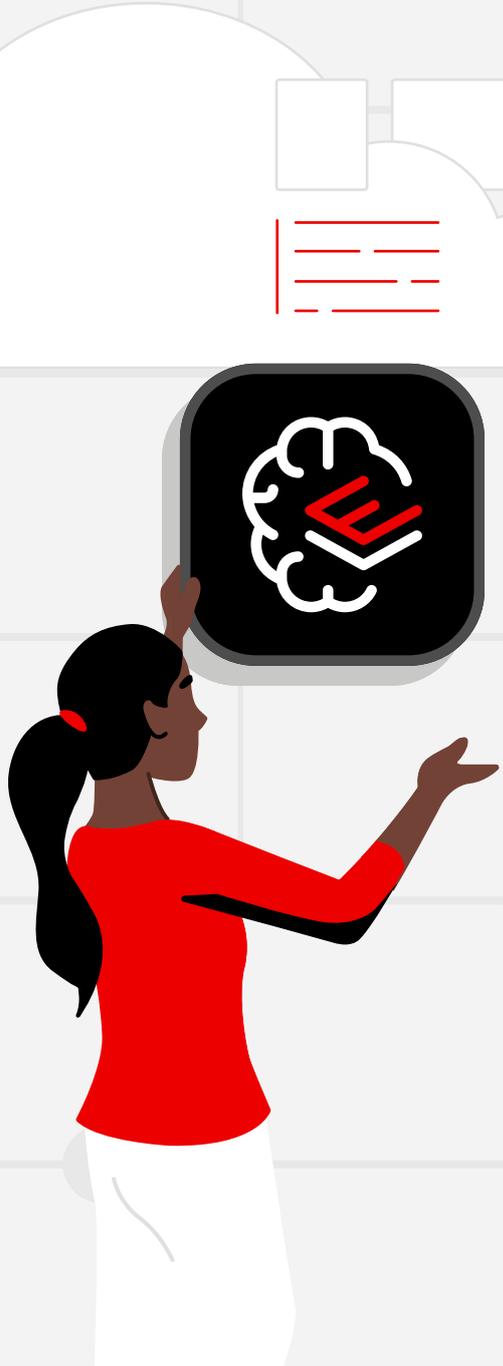
La gamme de produits Red Hat AI comprend Red Hat Enterprise Linux® AI, pour les environnements de serveurs Linux individuels, ainsi que Red Hat OpenShift® AI, pour les environnements de plateformes Kubernetes distribués. En outre, elle intègre les pratiques MLOps (Machine Learning Operations). Les deux solutions s'appuient sur des technologies et des modèles Open Source. Ainsi, les entreprises sont en mesure de rester à l'avant-garde de l'innovation en matière d'IA, d'accélérer le rythme des découvertes et de démocratiser l'accès aux outils et technologies de pointe.

Notre vaste écosystème de partenaires peut également enrichir vos capacités en matière d'IA. Connue pour avoir démocratisé le GPU et actrice majeure de l'IA, la société NVIDIA continue de collaborer avec Red Hat pour libérer le potentiel de cette technologie en proposant une plateforme d'entreprise complète optimisée pour les charges de travail d'IA. Elle permet aux clients professionnels de bénéficier de systèmes informatiques plus rapides pour les applications d'IA et de calcul haute performance.

**« Red Hat et NVIDIA collaborent étroitement depuis de nombreuses années, et Red Hat Enterprise Linux AI témoigne de notre volonté commune d'offrir des outils et des logiciels pour tous les systèmes aux équipes de développement et de recherche qui travaillent sur la prochaine génération de technologies et d'applications d'IA<sup>3</sup>. »**

**Justin Boitano**, vice-président de la division Enterprise Products, NVIDIA

<sup>3</sup> Communiqué de presse de Red Hat, « [Red Hat Delivers Accessible, Open Source Generative AI Innovation with Red Hat Enterprise Linux AI](#) », 7 mai 2024



## Présentation de Red Hat Enterprise Linux AI

La solution Red Hat Enterprise Linux AI comprend quatre composants principaux :

1

### Modèles Granite ouverts

La solution Red Hat Enterprise Linux AI inclut des modèles Granite Open Source et que nous prenons entièrement en charge. Ces modèles flexibles permettent de créer des modèles de langage personnalisés et de les utiliser de manière publique ou privée.

2

### Alignement de modèles avec InstructLab

InstructLab est un projet Open Source que nous menons avec IBM. Il personnalise les modèles d'IA avec des connaissances spécifiques et génère des données synthétiques destinées à l'entraînement des modèles. Grâce à cet outil en ligne de commande intégré à un référentiel Git, les utilisateurs peuvent facilement ajouter des compétences aux modèles et les entraîner.

3

### Version de démarrage optimisée de Red Hat Enterprise Linux pour les modèles Granite et InstructLab

Les modèles Granite et les outils InstructLab s'exécutent sur une image spéciale de Red Hat Enterprise Linux optimisée pour l'IA et compatible avec quasiment tous les équipements et environnements cloud. Cette configuration améliore les performances avec les GPU haut de gamme, ce qui accélère l'entraînement et le déploiement des modèles.

4

### Assistance pour les entreprises et protection de la propriété intellectuelle

Les souscriptions Red Hat Enterprise Linux AI incluent un service d'assistance pour les entreprises, un cycle de vie complet du produit à partir du modèle et du logiciel Granite 7B, ainsi qu'une clause d'indemnisation par Red Hat en cas de violation de la propriété intellectuelle.

## Red Hat Enterprise Linux AI donne vie aux applications d'IA générative

La solution Red Hat Enterprise Linux AI fournit aux entreprises qui débutent avec l'IA générative des LLM prêts à l'emploi ainsi que des modèles de langage de code dans un même environnement de développement de serveur et d'inférence.

Elles bénéficient ainsi d'un environnement unifié incluant des modèles et des outils qui leur permettent de se lancer plus facilement avec l'IA générative et de personnaliser des modèles à l'aide de leurs données métier, sans avoir besoin de compétences poussées en matière d'IA ni d'une infrastructure étendue.

La solution Red Hat Enterprise Linux AI, que nous prenons en charge et indemnisons entièrement, réduit les risques. Elle fournit une approche de l'IA simplifiée, plus accessible pour les équipes de développement et les spécialistes du domaine, qui peuvent ainsi collaborer et accélérer la génération de valeur pour l'entreprise.

## Les avantages de Red Hat Enterprise Linux AI

### Grands modèles de langage pour les entreprises

Les LLM Granite d'IBM, que nous prenons en charge et indemnisons entièrement, sont couverts par la licence Open Source Apache-2.0.

### Collaboration communautaire

InstructLab simplifie l'expérimentation de modèles d'IA générative et le réglage de l'alignement.

### Évolutivité cloud-native

Le mode image de Red Hat Enterprise Linux vous permet de gérer votre plateforme d'IA en tant qu'image de conteneur, rationalisant ainsi votre approche de la mise à l'échelle.

### Outils d'accélération et d'IA

Les accélérateurs matériels Open Source ainsi que les fonctions d'apprentissage profond optimisées permettent d'obtenir des résultats plus rapides.

## Mise à l'échelle avec Red Hat OpenShift AI

Red Hat OpenShift AI fournit une plateforme MLOps intégrée qui prend en charge la création, l'entraînement, le réglage, le déploiement et la surveillance à grande échelle des applications basées sur l'IA ainsi que des modèles de fondation et d'analyse prédictive, dans les environnements de cloud hybride.

Red Hat OpenShift AI s'appuie sur la plateforme Red Hat OpenShift pour offrir une expérience cohérente, rationalisée et automatisée en matière de gestion des charges de travail et des exigences de performance des projets d'IA/AA. Les pratiques MLOps peuvent aider les entreprises à réagir rapidement face aux innovations de l'IA et à accélérer la mise en production d'applications basées sur l'IA.

### Expérimentez dans le sandbox Red Hat OpenShift AI

Basée sur des composants issus de la plateforme Open Data Hub et d'autres projets Open Source, la solution Red Hat OpenShift AI offre aux équipes de science des données et de développement une plateforme d'IA/AA hybride et ouverte, optimisée pour extraire des renseignements à partir de données et pour créer des applications reposant sur l'IA.

[Essayer la solution dans notre sandbox de développement >](#)

## Les avantages de Red Hat OpenShift AI

### Adaptation de la mise à disposition des modèles

Les modèles peuvent être déployés pour être intégrés aux applications intelligentes sur site, dans le cloud public ou en périphérie du réseau. Il est possible de les recréer, de les redéployer et de les surveiller en fonction des modifications apportées au notebook source.

### Réduction du temps consacré à la gestion de l'infrastructure d'IA

Offrez à vos équipes un accès à la demande aux ressources. Elles pourront ainsi se concentrer sur l'exploration des données et la création d'applications qui apportent une valeur ajoutée à votre entreprise.

### Outils d'IA/AA testés et pris en charge

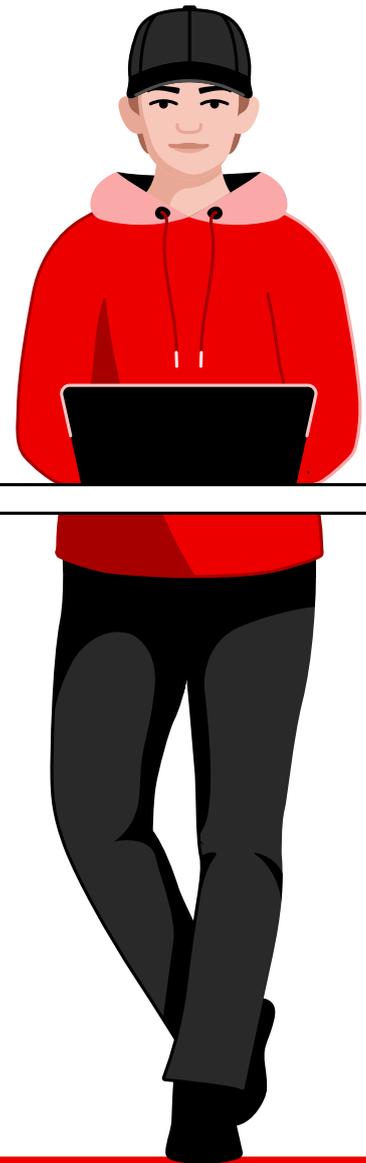
Nos équipes s'occupent du suivi, de l'intégration, des tests et de la prise en charge des outils courants et du déploiement de modèles d'IA/AA sur la plateforme d'applications Red Hat OpenShift.

### Flexibilité dans le cloud hybride

Disponible sous forme de logiciel autogéré ou de service cloud entièrement géré et exécuté sur Red Hat OpenShift, Red Hat OpenShift AI offre une plateforme flexible et axée sur la sécurité qui vous permet de développer et déployer vos modèles dans l'environnement de votre choix : sur site, dans le cloud public ou en périphérie du réseau.

### Exploitation selon nos meilleures pratiques

Grâce à nos services de consulting, vous pouvez installer, configurer et utiliser la solution Red Hat OpenShift AI pour exploiter tout son potentiel. Que vous souhaitiez réaliser une première expérience avec Red Hat OpenShift AI ou que vous cherchiez des conseils pour construire votre base MLOps, l'équipe de consulting Red Hat vous apportera son soutien et vous accompagnera tout au long de votre projet.



Avec notre gamme complète de technologies, notre expérience éprouvée et nos partenariats stratégiques, nous pouvons vous aider à atteindre vos objectifs en matière d'IA générative. Vous bénéficiez ainsi d'une base de développement et de déploiement pour les modèles et les applications d'IA générative, ainsi que de services et de formations pour accélérer leur adoption.

En savoir plus

# Envie de passer à l'étape suivante de votre parcours d'adoption ?

Accélérez l'adoption de l'IA avec notre stratégie de cloud hybride ouvert et exécutez vos applications en toute flexibilité, là où vous en avez besoin.

Profitez de nos spécialistes, de notre service de consulting et de nos formations pour lancer vos projets d'IA/AA et aider votre entreprise à atteindre ses objectifs en matière d'IA.

Découvrir les services pour l'IA/AA :

[red.ht/aiml-consulting](https://red.ht/aiml-consulting)

Organiser une session de découverte gratuite :

[redhat.com/consulting](https://redhat.com/consulting)

## En savoir plus sur Red Hat AI



[En savoir plus sur Red Hat Enterprise Linux AI](#)



[Découvrir la mise à l'échelle avec Red Hat OpenShift AI](#)