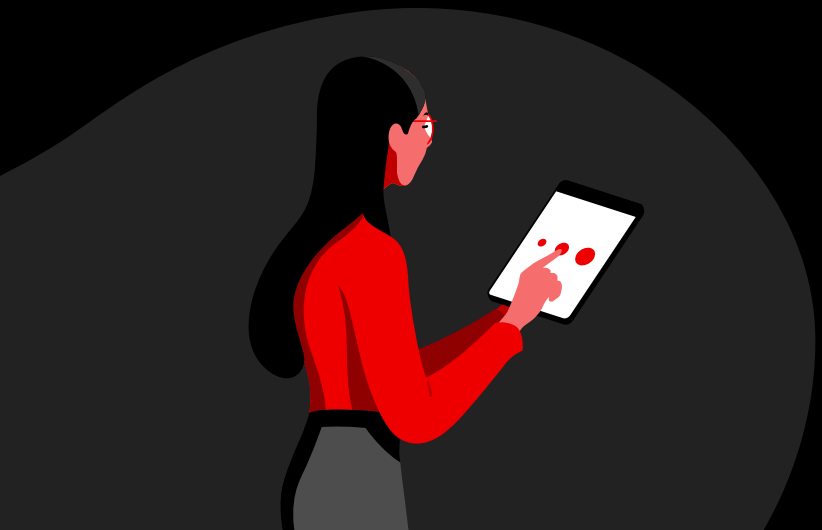
Red Hat


Éléments importants pour la

# création d'un environnement d'IA/AA prêt pour la production

# Sommaire



# Tirez davantage de valeur métier de vos données

D'ici 2026, la quantité de données créées devrait dépasser 221 000 exaoctets[1]. Dans un monde numérique, vos données peuvent représenter un immense avantage compétitif, mais il ne suffit pas de les recueillir : pour vous démarquer, vous devez les utiliser à bon escient.

Les technologies d'intelligence artificielle (IA), d'apprentissage automatique (AA) et d'apprentissage profond utilisent des données pour produire des informations métier, automatiser des tâches et enrichir les capacités des systèmes. Ces technologies ont le potentiel de transformer une entreprise à tous les niveaux, des clients et du personnel aux équipes de développement et d'exploitation. L'intégration de l'IA/AA à vos applications peut apporter des bénéfices mesurables à votre entreprise :

- Augmentation de la satisfaction des clients
- Enrichissement de l'offre de services numériques
- Optimisation des services métier existants
- Automatisation de l'exploitation métier
- Augmentation du chiffre d'affaires
- Amélioration de la prise de décisions
- Hausse de l'efficacité et baisse des coûts

## Technologies essentielles

Ce livre numérique présente plusieurs technologies qui permettent des analyses de données vraiment utiles :

- L'**intelligence artificielle** implique l'utilisation de machines qui imitent un comportement humain pour effectuer des tâches nécessitant normalement une intervention humaine.
- L'**apprentissage automatique** est une sous-catégorie de l'IA qui utilise des algorithmes et des modèles statistiques pour effectuer des tâches sans instructions explicites.
- L'**apprentissage profond** est une sous-catégorie de l'AA qui utilise des couches afin d'extraire progressivement des fonctions complexes à partir de données brutes, à la manière d'un cerveau humain. Par exemple, l'IA générative peut se baser sur des modèles d'apprentissage profond entraînés pour créer du texte, des images et d'autres contenus de qualité.
- Le **MLOps** englobe les outils, plateformes et processus nécessaires pour créer, entraîner, déployer, surveiller et améliorer en continu les modèles d'IA/AA destinés aux applications cloud-native.

1 *Livre blanc d'IDC, commissionné par Dell Technologies et NVIDIA, « **High Data Growth and Modern Applications Drive New Storage Requirements in Digitally Transformed Enterprises** », document n° US49359722, juillet 2022*

# Cas d'utilisation de l'IA/AA par secteur

Dans tous les secteurs, les technologies d'IA/AA peuvent aider les entreprises à atteindre plus rapidement leurs objectifs.

## Services financiers

- Personnalisation des services et offres pour les clients
- Amélioration de l'analyse des risques
- Détection des fraudes et du blanchiment d'argent

## Télécommunications

- Informations sur les comportements des clients
- Amélioration des expériences client
- Optimisation des performances du réseau 5G

## Distribution

- Optimisation des chaînes d'approvisionnement et de la gestion des inventaires
- Amélioration de la qualité des informations et des expériences client

## Industrie automobile

- Soutien aux technologies de conduite autonome
- Prévision des besoins d'entretien des équipements
- Amélioration des chaînes d'approvisionnement

## Santé

- Augmentation de l'efficacité des hôpitaux et des cliniques
- Amélioration de la vitesse et de la précision des diagnostics
- Amélioration des résultats pour les patients

## Énergie

- Optimisation des interventions et de la maintenance sur le terrain
- Renforcement de la sécurité du personnel
- Rationalisation du commerce de l'énergie

## Fabrication

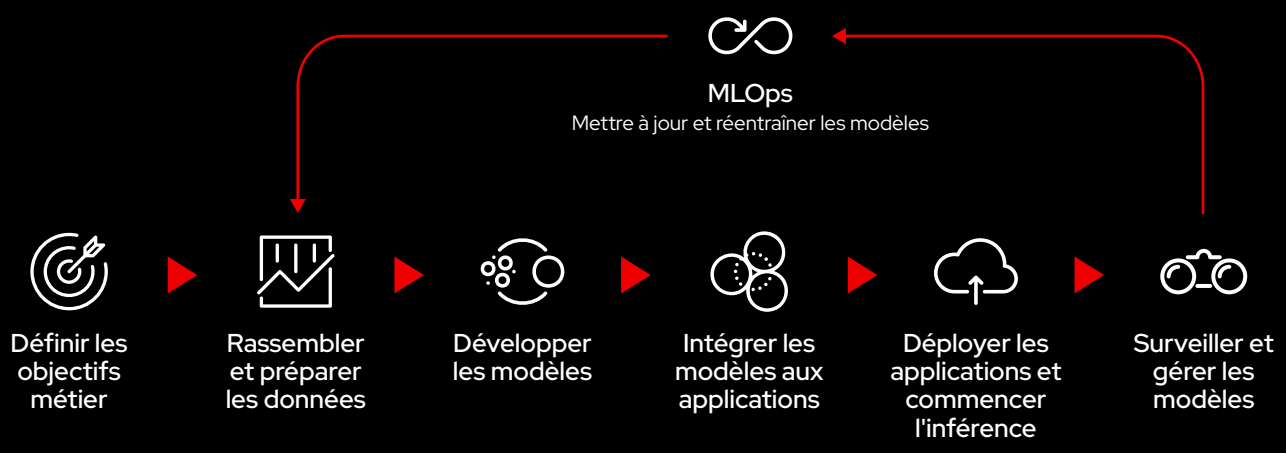
- Prévision des pannes d'équipements
- Réalisation d'entretiens préventifs
- Amélioration de la sécurité dans l'usine

# Créez un environnement d'IA/AA prêt pour la production

Le déploiement de l'IA/AA en production est un processus itératif qui va plus loin que la simple création de modèles d'IA/AA. Voici les principales étapes du cycle de vie de l'IA/AA :

1. Fixez des objectifs métier pour votre projet d'IA/AA et partagez-les avec toutes les parties prenantes.
2. Rassemblez et préparez les données nécessaires pour votre projet d'IA/AA.
3. Développez des modèles en fonction de vos objectifs.
4. Déployez des modèles dans votre processus de développement d'applications.
5. Mettez en œuvre des applications intelligentes basées sur l'AA et commencez l'inférence des modèles.
6. Surveillez et gérez la précision des modèles au fil du temps.

## Cycle de vie de l'IA/AA et MLOps



## Défis liés au déploiement de l'IA/AA

Lors de la création d'un environnement d'IA/AA, les entreprises peuvent rencontrer plusieurs défis :

- **Manque de talents :** parce que les spécialistes de l'IA/AA se font rares, il est difficile de trouver et fidéliser du personnel compétent pour constituer des équipes de science des données, d'ingénierie, d'ingénierie de l'AA ou encore de développement logiciel.
- **Absence de données prêtes à l'emploi :** les entreprises recueillent de grandes quantités de données et doivent identifier, préparer et protéger les bonnes données pour chaque projet d'IA/AA.
- **Disparité des équipes et des technologies :** si l'exploitation et l'infrastructure sont lentes, déconnectées et basées sur des processus manuels, les équipes peuvent moins facilement collaborer et le déploiement de l'IA/AA peut être ralenti.
- **Retard dans la disponibilité des ressources :** une distribution lente de l'infrastructure et des outils nuit au développement, à l'intégration et au déploiement de modèles dans les applications.

Vous pouvez cependant surmonter ces défis en appliquant des approches cloud-native de développement d'applications au cycle de vie de l'IA/AA.

**Avec une architecture ouverte et adaptable, vous pouvez adopter l'IA/AA et les pratiques MLOps plus efficacement afin d'atteindre vos objectifs métier.**

Une architecture d'IA/AA prête pour la production nécessite plusieurs technologies et capacités clés :

- **Des outils d'IA/AA et MLOps** qui permettent aux équipes de science des données, d'ingénierie de l'AA et de développement d'applications de créer, déployer et gérer des modèles et applications d'AA
- Une **plateforme cloud** qui donne aux équipes d'ingénierie des données, de science des données, d'ingénierie de l'AA et de développement d'applications un accès aux ressources dont elles ont besoin pour travailler rapidement
- **Des accélérateurs de calcul, stockage et réseau** qui raccourcissent les délais de préparation des données, le développement des modèles et les tâches d'inférence
- **Des points de terminaison d'infrastructure** qui fournissent des ressources pour les environnements de cloud privé, public et hybride, sur site, virtuels et d'edge computing pour toutes les étapes de l'exploitation de l'IA/AA
- **Des déploiements d'edge computing** (facultatifs) qui donnent accès à de nombreuses données issues des appareils et des capteurs afin d'entraîner les modèles et d'obtenir des informations en temps réel

Ce livre numérique présente les éléments essentiels à prendre en compte pour la création d'une architecture d'IA/AA efficace.

# Conteneurs

Un **conteneur** est une unité de base d'un logiciel qui contient une application avec toutes ses dépendances.

D'un côté, les équipes de science des données, d'ingénierie de l'AA et de développement d'applications ont besoin d'accéder aux outils et ressources de leur choix pour optimiser leur productivité. De l'autre, les équipes d'exploitation doivent s'assurer que les ressources sont à jour, conformes et utilisées de manière sécurisée. Les conteneurs simplifient les processus de création d'applications et permettent leur déploiement dans différents environnements, sans modifications. Ils permettent de déployer une large sélection d'outils d'IA/AA dans des environnements hybrides et ce, de manière cohérente. Les équipes peuvent modifier de manière itérative et partager des images de conteneurs grâce au système de contrôle de version qui suit les changements entre les versions pour plus de transparence. En parallèle, les fonctionnalités d'isolation des processus et de contrôle des ressources améliorent la protection contre les menaces.

### Recommandations relatives aux solutions de conteneurs

Cherchez une plateforme de conteneurs robuste et hautement disponible qui inclut des fonctions de sécurité intégrées et simplifie le déploiement, la gestion et les déplacements de conteneurs dans votre environnement. Choisissez une plateforme Open Source compatible avec un large éventail de technologies pour gagner en flexibilité et en choix.

# Orchestration des conteneurs

L'orchestration des conteneurs consiste à gérer la création, le déploiement et le cycle de vie des conteneurs dans l'ensemble de votre environnement.

Une fois que vous avez adopté les conteneurs, vous avez besoin d'un moyen efficace pour les déployer, les gérer et les faire évoluer. Avec un outil d'orchestration des conteneurs, vous pouvez administrer le cycle de vie de vos conteneurs de manière cohérente. Ce type d'outil centralise l'accès aux ressources de calcul, stockage et mise en réseau dans les environnements sur site, cloud et d'edge computing. Ces outils fournissent également des fonctions de planification unifiée des charges de travail, de contrôle multi-client et d'application de quotas.

### Recommandations relatives à l'orchestration des conteneurs

Choisissez un outil d'orchestration basé sur **Kubernetes** pour tirer parti de cette technologie Open Source de pointe.

# Gestion du cycle de vie des applications

La gestion du cycle de vie des applications couvre le déploiement, la mise à l'échelle et l'administration des applications exécutées dans des conteneurs.

Par nature, les environnements d'IA/AA sont complexes. Vous pouvez utiliser des composants de gestion du cycle de vie des applications conteneurisées qui fonctionnent avec votre outil d'orchestration des conteneurs pour administrer directement les applications conteneurisées, notamment les outils de développement de l'IA/AA. Les équipes d'exploitation peuvent automatiser des tâches courantes de gestion du cycle de vie telles que la configuration, le provisionnement et les mises à jour pour gagner en efficacité, en rapidité et en précision. Les équipes de science des données, d'ingénierie de l'AA et de développement d'applications peuvent utiliser les outils et applications d'un catalogue de services préapprouvés, sans faire appel aux équipes d'exploitation. L'automatisation libère également les équipes des tâches fastidieuses et leur permet de se concentrer sur des activités stratégiques plus intéressantes.

### Recommandations relatives à la gestion des applications

Choisissez des outils de gestion du cycle de vie des applications conteneurisées qui incluent des fonctions d'automatisation faciles à utiliser et qui fonctionnent avec vos outils d'IA/AA préférés. Parmi les outils les plus utilisés figurent les **opérateurs Kubernetes** et les **charts Helm**.

# Pratiques MLOps

Les pratiques MLOps rassemblent les outils, plateformes et processus nécessaires pour exploiter l'IA/AA à grande échelle.

Les entreprises doivent développer et déployer rapidement et efficacement des modèles d'IA/AA ainsi que les applications qui les utilisent. La collaboration entre les équipes joue ici un rôle essentiel. Tout comme le DevOps, les approches MLOps favorisent cette collaboration entre les équipes d'IA/AA, de développement d'applications et d'exploitation informatique, dans le but d'accélérer la création, l'entraînement, le déploiement et la gestion des modèles d'AA et des applications basées sur l'AA. L'automatisation, souvent sous la forme de pipelines d'**intégration et de distribution continues (CI/CD)**, permet des changements rapides, progressifs et itératifs qui accélèrent les cycles de développement des applications et des modèles.

### Meilleures pratiques MLOps

En plus des technologies, les pratiques MLOps englobent les équipes et les processus. Appliquez des **pratiques MLOps** à l'ensemble du cycle de vie de l'IA/AA. Utilisez l'automatisation sur vos plateformes et dans vos outils, ainsi que des technologies Open Source comme Argo, Kubeflow, Tekton et Jenkins pour créer des pipelines et workflows CI/CD.

# Plateforme de cloud hybride

Une plateforme de cloud hybride fournit une base pour le développement, le déploiement et la gestion d'applications intelligentes et de modèles dans les environnements sur site, cloud et d'edge computing.

Les modèles d'IA/AA et les applications intelligentes nécessitent une infrastructure pour leur développement et leur déploiement. Avec une plateforme de cloud hybride cohérente, vous pouvez développer, tester, déployer et gérer vos modèles et applications de la même manière dans toutes les parties de votre infrastructure. Ce type de plateforme offre suffisamment de portabilité, d'évolutivité et de flexibilité pour provisionner des environnements d'IA/AA à la demande. Cette plateforme peut également fournir des capacités en libre-service pour accélérer la distribution des ressources tout en préservant leur contrôle. Enfin, une plateforme cohérente fournit une base pour intégrer les solutions technologiques des vendeurs tiers et des communautés Open Source, ainsi que tous les outils développés sur mesure de votre choix.

### Recommandations relatives aux plateformes de cloud hybride

Choisissez une plateforme centrée sur la sécurité qui prend en charge l'accélération du matériel, qui s'appuie sur un vaste écosystème d'outils de développement d'applications et d'IA/AA et qui intègre des capacités de gestion et DevOps. Les plateformes Open Source peuvent offrir plus de possibilités d'intégration et de flexibilité.

# Déploiements d'edge computing

**Les déploiements d'edge computing** sont des environnements décentralisés dans lesquels les appareils recueillent des données et exécutent des fonctions sur site, en dehors du datacenter central.

L'edge computing permet de fournir des informations et des expériences à l'endroit et au moment où elles sont utiles. Souvent, les capteurs et les appareils génèrent un grand volume de données qui peuvent servir à l'entraînement des modèles et à l'inférence de l'exécution dans les workflows d'IA/AA. Le transfert de ces données vers un cloud central en temps réel peut se révéler cher et complexe. Il est par exemple possible d'utiliser des algorithmes de reconnaissance d'image qui s'exécutent plus efficacement près de la source des données, pour supprimer le besoin de transférer de grandes quantités de données à traiter.

### Meilleures pratiques en matière d'edge computing

Les capacités d'évolutivité, de connectivité et de gestion des appareils sont essentielles pour les déploiements d'edge computing. Optez pour des solutions que vous pouvez gérer avec les mêmes outils et processus que ceux utilisés pour votre datacenter et votre infrastructure cloud. Choisissez une plateforme capable de gérer l'interruption des communications et la déconnexion des environnements. Recherchez également des solutions offrant davantage de flexibilité et de possibilités de personnalisation, qui prennent en charge de nombreux appareils et équipements.

# Déployez une plateforme de base ouverte et flexible

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/AA. Nous vous proposons une base solide pour la création d'environnements d'IA/AA prêts pour la production, ainsi que des services et des formations pour accélérer l'adoption de notre solution.

**Red Hat® OpenShift®** est une plateforme d'applications unifiée pour les entreprises, conçue pour l'innovation cloud-native. Cette solution offre à vos équipes la rapidité et la flexibilité dont elles ont besoin pour réussir, avec les ressources de calcul à la demande, la prise en charge de l'accélération du matériel et du processeur graphique (GPU), ainsi que la cohérence dans les environnements sur site, de cloud public et d'edge computing. Vous pouvez par exemple créer une plateforme MLOps en libre-service pour les équipes de science des données, d'ingénierie des données et de développement afin de produire rapidement des modèles, de les incorporer aux applications et de réaliser des tâches d'inférence. Les fonctions de collaboration permettent aux équipes de créer et partager des résultats de modélisation dans des conteneurs avec leurs homologues et les développeurs de manière cohérente.

La gamme Red Hat OpenShift AI regroupe des produits pour l'entraînement, la distribution, la surveillance et la gestion du cycle de vie des modèles et applications d'IA/AA. Elle comprend notamment le service **Red Hat OpenShift Data Science**, qui offre aux équipes de science des données et de développement une puissante plateforme d'IA/AA pour rassembler les informations et créer des applications intelligentes. Les équipes peuvent passer de l'expérimentation à la production dans un environnement cohérent et collaboratif qui intègre les offres des principaux partenaires certifiés, notamment NVIDIA, Intel, Starburst, Anaconda, IBM et Pachyderm.

La gamme **Red Hat Application Services** inclut des outils et services qui permettent de créer un environnement unifié pour le développement, la distribution, l'intégration et l'automatisation des applications. Les services d'intégration des données offrent la possibilité de construire des pipelines de données efficaces, tandis que les services d'exécution simplifient le développement des applications. Les outils et services d'automatisation des processus peuvent accéder à des applications intelligentes ainsi qu'à des modèles d'AA pour automatiser la prise de décisions métier.

Enfin, les produits de plateforme Red Hat, notamment **Red Hat Enterprise Linux®**, **Red Hat OpenStack® Platform** et **Red Hat OpenShift Platform Plus**, offrent une infrastructure logicielle évolutive.

## Créer avec la communauté

Nous contribuons activement aux communautés Open Source **Kubeflow** et **Open Data Hub**. Le projet communautaire Open Data Hub fournit un modèle pour intégrer les outils d'IA/AA Open Source courants à un environnement OpenShift. Des outils d'analyses des données et d'AA courants, comme Ray, Ceph®, Apache Kafka, Kubeflow, TensorFlow et Jupyter Notebooks, sont intégrés à l'architecture de référence.

## Gagnez en flexibilité avec un écosystème de partenaires certifiés

L'**écosystème de partenaires certifiés Red Hat** permet d'intégrer des outils d'IA/AA, d'analyses des données, de gestion, de stockage, de sécurité et de développement à cette architecture. Chez Red Hat, nous travaillons en étroite collaboration avec nos partenaires afin de certifier leurs logiciels sur nos plateformes, pour plus de facilité de gestion, une sécurité renforcée et une meilleure prise en charge. En outre, de nombreux partenaires proposent des **opérateurs Red Hat OpenShift certifiés** qui simplifient la gestion du cycle de vie des logiciels.

## Choisissez les produits et technologies que vous préférez

Grâce à notre écosystème de partenaires d'IA/AA certifiés, vous pouvez intégrer des produits et technologies couramment utilisés à votre environnement.


**NVIDIA** et Red Hat proposent des solutions pour accélérer la distribution des applications intelligentes basées sur l'IA dans tous les environnements. L'association des solutions **NVIDIA AI Enterprise et Red Hat OpenShift** permet d'accéder à une gamme complète de logiciels d'IA et d'analyses des données, optimisés et cloud-native. Les systèmes Red Hat Enterprise Linux, Red Hat OpenShift et NVIDIA DGX facilitent la gestion informatique de l'infrastructure d'IA. Enfin, la solution **NVIDIA GPU Operator** automatise la gestion de tous les composants logiciels NVIDIA qui interviennent dans le provisionnement des GPU.


**Starburst** et Red Hat proposent des solutions pour exploiter les sources de données distribuées. La solution **Starburst Enterprise** s'utilise avec Red Hat OpenShift pour accélérer l'analyse des données sur plusieurs plateformes, Cette association offre de nombreuses capacités : automatisation à l'échelle de l'entreprise, haute disponibilité, élasticité et surveillance. Grâce à cette solution, les entreprises peuvent moderniser les données, exécuter des charges de travail ETL (extraction, transformation et chargement), réaliser des analyses interactives des données et alimenter les outils d'informatique décisionnelle.


**Intel** et Red Hat collaborent pour offrir une infrastructure logicielle et des plateformes standard qui favorisent l'agilité et la flexibilité du datacenter. La distribution Intel du **kit d'outils OpenVINO** optimise et convertit les modèles d'apprentissage profond en moteurs d'inférence à hautes performances qui peuvent s'adapter automatiquement à des milliers de nœuds sur Red Hat OpenShift. Le **kit d'outils Intel AI Analytics, optimisé par oneAPI**, offre un ensemble complet d'outils logiciels d'IA interopérables pour accélérer et mettre à l'échelle les workflows d'AA.


**SAS** et Red Hat collaborent afin de développer des technologies de cloud hybride ouvert et des capacités d'analyse qui permettent d'obtenir des informations utiles et exploitables. Exécutée sur Red Hat OpenShift, la plateforme de cloud hybride **SAS Viya** donne accès aux principales applications d'analyses, d'IA et d'AA pour permettre aux entreprises de créer des applications puis de les déployer dans tous les environnements. Grâce à une gestion cohérente des infrastructures, les équipes sont unies et peuvent mieux collaborer. Cette plateforme unifiée permet aux entreprises de développer et déployer des modèles avec les interfaces, langages et infrastructures de leur choix.

# Découvrez les réussites de clients


Avec l'aide de l'équipe de consulting Red Hat, **Banco Galicia** a créé une plateforme de traitement du langage naturel (TLN) intelligente et basée sur l'IA qui repose sur les solutions Red Hat OpenShift, Red Hat Integration et Red Hat Single Sign-On (SSO).

La banque a réduit le délai d'intégration des clients professionnels,

**de 20 jours à quelques minutes**

tout en atteignant un taux de précision de 90 % pour les analyses des données.

Lire le **témoignage client**

## Département des Anciens combattants des États-Unis

Le groupe **Team Guidehouse du département des Anciens combattants des États-Unis** a déployé Red Hat OpenShift et Red Hat OpenShift Data Science afin d'utiliser les techniques d'AA dans un prototype de solution destiné à prévenir le suicide chez les anciens combattants.

**Lauréat de la phase 2**

du défi Mission Daybreak

Lire l'**article de blog**


**Nippon Telegraph and Telephone East Corporation (NTT East)** a créé un service d'analyses des données d'edge computing avec Red Hat OpenShift.

« [...] Red Hat OpenShift nous a permis de développer et d'exécuter de manière stable des services vidéo novateurs basés sur l'IA, en collaboration avec l'équipe de développement de l'IA. »

**Masashi Toyama**
Responsable des technologies de l'infrastructure de serveur pour le service d'ingénierie du serveur cloud, division de promotion avancée – Siège commercial du réseau de NTT East

Lire le **témoignage client**


L'**université de Boston** utilise Red Hat OpenShift Data Science comme plateforme principale pour les cours de sciences de l'informatique et de systèmes d'ingénierie informatique.

« Mes étudiants bénéficient ainsi d'une expérience Linux riche et complète, aussi exhaustive qu'accessible. Je peux même l'intégrer à mes supports et méthodes d'enseignement. »

**Jonathan Appavoo**
Professeur associé à l'université de Boston

Lire l'**article de blog**

# Vous souhaitez vous lancer dans l'IA/AA ?

## Les technologies d'IA/AA et MLOps transforment presque tous les aspects d'une entreprise.

Nous pouvons vous aider à créer un environnement d'IA/AA prêt pour la production qui accélère le développement et la distribution des applications intelligentes afin de soutenir vos objectifs métier.

**Red Hat OpenShift AI**

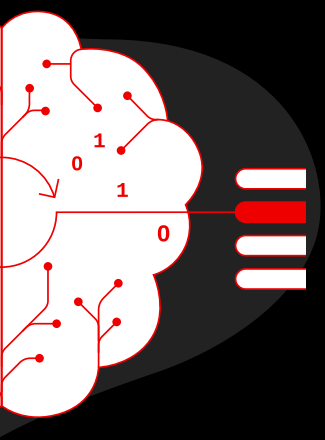
Découvrez comment Red Hat OpenShift peut accélérer les workflows d'IA/AA et la distribution des applications intelligentes basées sur l'IA : red.ht/OpenShiftAI

**Essayer gratuitement**

**Red Hat OpenShift Data Science**

Découvrez comment Red Hat OpenShift Data Science peut faciliter la mise en place des meilleures pratiques MLOps : red.ht/datascience

**Essayer gratuitement**




## Lancez-vous immédiatement avec les services de consulting Red Hat

Travaillez avec nos spécialistes pour démarrer sans attendre vos projets d'IA/AA. Nous proposons des services de consulting et de formation pour aider votre entreprise à adopter l'IA/AA plus rapidement.

- Découvrez nos services pour l'IA/AA : **red.ht/aiml-consulting**
- Organisez une session de découverte gratuite : **redhat.com/consulting**



479615_0823_KVM

Red Hat