4 principes clés à prendre en compte pour mettre en œuvre des technologies d'IA

Les chefs d'entreprise qui souhaitent mettre en œuvre des technologies d'intelligence artificielle (IA) ou développer leur utilisation partagent un certain nombre de questions et de préoccupations. Ils cherchent notamment à comprendre comment utiliser l'IA pour améliorer l'efficacité de l'exploitation et développer leur activité tout en préservant la stabilité et en réduisant les risques. Voici quatre principes à prendre en compte, quelle que soit la situation de votre entreprise en matière de mise en œuvre de l'IA.

1 Choisir une plateforme flexible

Que vous commenciez tout juste à déployer l'IA ou que vous cherchiez à développer les cas d'utilisation existants et généraliser sa mise en œuvre, il est essentiel d'opter pour une plateforme flexible et évolutive qui s'adaptera aux déploiements de vos technologies d'IA.

Aux premières étapes de l'adoption, vous avez besoin d'une infrastructure ouverte et flexible qui permettra à vos équipes de créer et d'exécuter de petites charges de travail. La plateforme que vous utilisez doit prendre en charge l'entraînement des modèles à grande échelle pour suivre la hausse de l'utilisation de l'IA ainsi que le développement des charges de travail.

La stabilité de la plateforme permettra à vos équipes de science des données et de développement de créer des modèles à forte valeur ajoutée pour vos clients. Ces derniers auront alors l'assurance d'utiliser une plateforme fiable et axée sur la sécurité.

Si votre entreprise est déjà plus avancée en matière d'IA, vous avez besoin des technologies qui permettront d'exploiter pleinement son potentiel. Vous devrez peut-être investir de manière significative dans des processeurs graphiques (GPU) et autres accélérateurs de matériel.

Avec une plateforme ouverte et évolutive, vous pouvez :

- Fournir à votre entreprise une base solide pour créer et moderniser plus rapidement les applications basées sur l'IA
- Accélérer le déploiement des applications basées sur l'IA
- Renforcer votre compétitivité dans un paysage de l'IA qui évolue rapidement

Avec l'infrastructure adéquate, votre entreprise sera en mesure d'adopter toutes les futures innovations en matière d'IA.

2 Se préparer pour l'IA hybride

L'IA générative requiert plus de ressources de calcul et place les capacités de traitement en périphérie du réseau. Les cas d'utilisation de l'IA générative reposent sur des ressources sur site aussi bien que dans le cloud, mais certaines charges de travail, notamment celles concernées par les problèmes de conformité ou de gravité des données, doivent être exécutées exclusivement sur site. Avec l'IA hybride, votre équipe est en mesure d'exécuter des charges de travail à l'emplacement le plus pertinent, y compris au plus près du site où sont générées les données.

L'IA hybride présente un certain nombre d'avantages, notamment :

- **Flexibilité :** la possibilité de choisir l'environnement le plus adapté pour exécuter des charges de travail d'expérimentation et de production permet de réduire les coûts.
- **Accélération de la distribution :** la réalisation de l'inférence des petites charges de travail en périphérie du réseau accélère la distribution des informations destinées aux utilisateurs.
- **Réduction des risques :** l'exécution d'applications d'IA en périphérie permet de s'affranchir du transfert d'une grande quantité d'informations vers le centre du réseau, et donc de renforcer la protection des données.

Avec l'IA hybride, votre équipe informatique peut entraîner un modèle dans un datacenter central ou le cloud public, puis le distribuer aux points de terminaison, où il peut exécuter l'inférence et fournir des informations utiles et exploitables plus rapidement. Les charges de travail qui nécessitent davantage de ressources peuvent être réparties entre le cloud central et la périphérie afin d'adapter la puissance de traitement pour fournir rapidement des recommandations précises.

3 Dépasser le modèle initial

Si vos équipes informatiques peuvent probablement lancer un modèle d'IA en production sans trop d'efforts, la réussite de cette opération reste souvent conditionnée par la capacité de mise à l'échelle. Pour maximiser la valeur métier, vos équipes doivent être en mesure de déployer plusieurs versions de ce modèle, potentiellement des centaines de fois chaque jour.

Elles pourront plus facilement mettre à l'échelle les déploiements avec des processus MLOps (Machine Learning Operations). Tout comme l'approche DevOps, le modèle MLOps rassemble les équipes de développement, d'exploitation et de science des données pour accélérer le développement d'applications basées sur l'IA et réussir leur déploiement. Grâce à ce modèle, votre entreprise peut mettre à l'échelle plus facilement le développement d'applications, et donc offrir davantage de services aux clients plus rapidement.

Le MLOps peut aussi favoriser l'efficacité et la rentabilité du développement. En collaborant sur une plateforme commune, les équipes gagnent en efficacité et peuvent ainsi réduire les délais et les coûts, ce qui aura un effet direct sur les résultats de votre entreprise.

4 Accorder la priorité à la sécurité

Pour éviter d'être confrontés à des cyberattaques coûteuses, les chefs d'entreprise cherchent de plus en plus à protéger les données sensibles et à réduire les risques. L'arrivée de l'IA générative a en plus imposé d'agir rapidement puisqu'elle augmente le risque d'exposition des données confidentielles.

Il est essentiel de renforcer votre posture de sécurité pour éviter les menaces suivantes :

- **Empoisonnement de données :** attaque qui consiste à introduire des données corrompues ou malveillantes dans les modèles d'IA pour générer des informations imprécises ou des recommandations incorrectes.
- **Vol de modèle :** attaque qui consiste à inverser l'ingénierie d'un modèle d'apprentissage automatique et à extraire des données en faisant une copie.
- **Attaque par porte dérobée :** attaque qui consiste à dissimuler du code malveillant dans un modèle d'IA afin que le pirate puisse voler des informations.

Ces tactiques ainsi que d'autres mettent votre entreprise en danger en l'exposant à des risques financiers aussi bien que juridiques. Minimisez les vulnérabilités associées à l'IA en vous appuyant sur un fournisseur de plateformes d'applications qui a fait ses preuves et mis en place des processus de gestion des correctifs de sécurité et des versions.

En savoir plus

Accélérez votre stratégie d'IA. Découvrez les solutions de Red Hat pour 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.

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

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

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

fr.redhat.com