

Accélérer les décisions  
data-driven génératrices de  
performance business dans votre  
entreprise grâce à l'adoption du  
**MLOps**

Découvrez les enjeux de la gestion de la chaîne  
de production de modèles de machine learning et  
les avantages de l'adoption du MLOps pour son  
optimisation



# Introduction

## La vérité établie

Il est désormais établi qu'une maîtrise des données permet d'améliorer considérablement l'efficacité opérationnelle. Exemples :

- Maximiser la performance des campagnes marketing en utilisant des algorithmes pertinents pour délivrer le bon message au bon client et au bon moment.
- Optimiser le parcours client en identifiant les meilleures opportunités ou points de friction qui affectent l'engagement, la rétention ou encore le taux de conversion.

## Le paradoxe

Le déploiement des cas d'usages à fort impact business reste peu courant dans de nombreuses entreprises. Plusieurs études et rapports le démontrent :

- L'étude "**Data and Analytics Maturity Model**" de Gartner montre que seulement 30% des entreprises ont atteint un niveau élevé de maturité en matière de maîtrise des données.
- Le rapport "**The State of Data and Analytics**" de Forbes Insights montre que seulement 32% des entreprises ont réussi à déployer des cas d'utilisation à fort impact business.

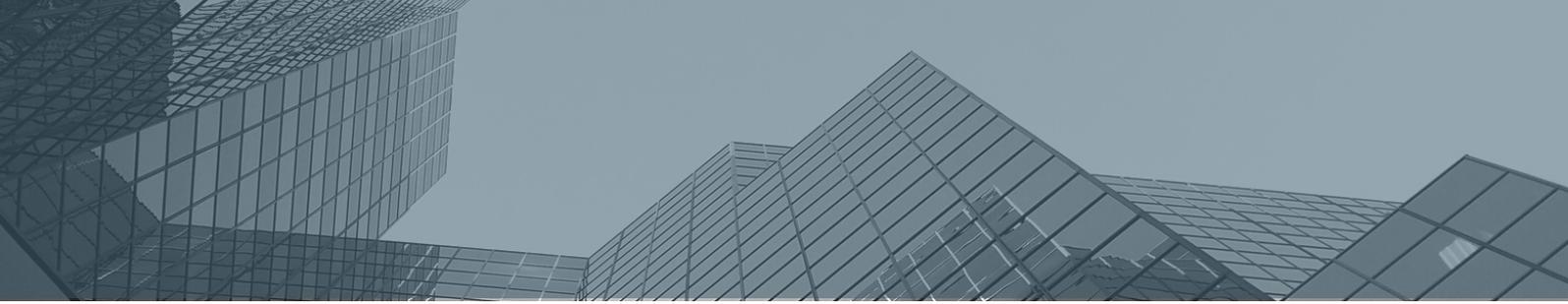
## Le diagnostic

Les causes sont multiples et peuvent inclure :

- **Le manque de vision stratégique** : certaines entreprises ne réalisent pas pleinement l'importance d'intégrer la donnée au sein de leurs processus décisionnels et peuvent ne pas inclure cela dans leur stratégie globale.
- **Les obstacles financiers** : les coûts liés à l'acquisition ou au développement des solutions data, ainsi que les dépenses d'opération, peuvent être perçus comme trop élevés par certaines entreprises et peuvent les empêcher de déployer des cas d'usage à fort impact business.
- **La complexité de la mise en place** : la difficulté à organiser la collaboration efficace entre les équipes data, IT et métier est un frein. En effet, il est essentiel de coordonner les efforts de ces différentes équipes pour construire, déployer et opérer des cas d'usage, mais cela peut s'avérer difficile en raison des divergences d'objectifs et de compétences.

En somme, les initiatives peinent souvent à passer de la phase de PoC / PoV à celle de mise en place à grande échelle, empêchant ainsi de tirer pleinement parti du potentiel.

C'est là que le MLOps entre en jeu. Il offre les possibilités de surmonter ces obstacles et accélérer ainsi la prise de décisions data-driven.



01

## **Les enjeux de la gestion de la chaîne de production de modèles de machine learning**

Les principaux défis.....P.4

Les conséquences d'une mauvaise gestion.....P.5

02

## **Les avantages de l'adoption du MLOps pour optimiser la gestion de la chaîne de production de modèles de machine learning**

La définition du MLOps.....P.6

Comment le MLOps peut aider à résoudre les défis ?..P.7

03

## **La mise en place une stratégie de MLOps**

Les étapes .....P.9

Les points clés à prendre en compte .....P.10

Les 3 erreurs à éviter .....P.11



# Les enjeux de la gestion de la chaîne de production de modèles de machine learning



## Les principaux défis

- **Complexité des modèles** : les modèles sont de plus en plus complexes, ce qui rend leur gestion plus difficile. Les modèles utilisent de plus en plus des architectures de réseaux de neurones profonds, des ensembles de données volumineux, des algorithmes de traitement de données avancés, etc. Il est donc nécessaire de disposer de compétences et de ressources pour gérer efficacement cette complexité.
- **Automatisation et standardisation** : il est nécessaire d'automatiser et de standardiser les différentes étapes de la chaîne de production pour gérer efficacement les modèles en production. Cela permet de réduire les coûts et de gagner en temps, mais cela nécessite également des investissements importants en termes de développement des briques logiciels et de processus associés.
- **Surveillance et maintenance** : les modèles en production doivent être surveillés et maintenus pour garantir leur qualité et leur performance. Cela nécessite des processus et des outils de surveillance et de maintenance adaptés, ainsi que des compétences pour les utiliser efficacement.
- **Gouvernance et conformité** : les modèles doivent respecter des normes éthiques pour éviter tout risque juridique ou préjudice pour les utilisateurs finaux. Il est important de s'assurer qu'ils ne comportent pas de biais, qui peuvent se produire lorsque les données d'entraînement sont inéquitables. Un exemple de biais dans un modèle de reconnaissance de la parole est lorsque le modèle est entraîné avec des enregistrements d'accent standard, il peut avoir des difficultés à comprendre les accents différents.
- **Collaboration inter-équipe** : les différents acteurs impliqués dans la mise en production de modèles doivent collaborer efficacement. Exemples :
  - Il est important que les Data Scientist et de Data Engineer utilisent les mêmes outils et plateformes pour éviter des re-développements inutiles qui peuvent causer des biais, des dysfonctionnements et des erreurs dans les modèles, compromettant ainsi leur fiabilité.
  - Il est important que les data scientists comprennent les besoins métier et que les équipes métier comprennent les limites et les possibilités des solutions proposées

# Les enjeux de la gestion de la chaîne de production de modèles de machine learning



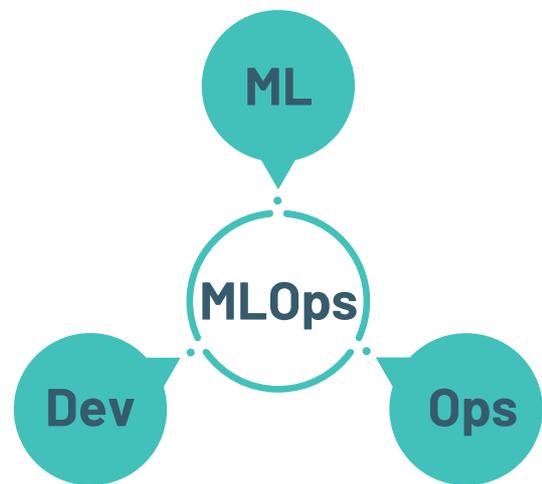
## Les conséquences d'une mauvaise gestion

- **Performances des modèles dégradées** : si les modèles ne sont pas gérés efficacement, ils peuvent présenter des performances dégradées, ce qui peut entraîner une mauvaise prise de décision avec des conséquences financières importantes.
- **Perte de confiance** : les modèles mal gérés peuvent entraîner une perte de confiance des utilisateurs finaux et des partenaires, qui peuvent ne plus vouloir les utiliser en raison de préoccupations liées à la qualité des résultats.
- **Conformité et gouvernance non respectée** : les modèles mal gérés peuvent également entraîner des problèmes juridiques, tels que des poursuites pour violation de la vie privée ou de la discrimination. Il est donc important de s'assurer que les modèles sont gérés de manière responsable et éthique, et de prendre les mesures nécessaires pour éviter tout risque.
- **Coûts élevés** : si les modèles ne sont pas gérés efficacement, ils peuvent entraîner des coûts élevés. Par exemple, le développement de modèles peut être coûteux, surtout s'ils nécessitent des données de qualité supérieure ou des outils de pointe. De plus, une maintenance régulière est nécessaire pour s'assurer que les modèles fonctionnent correctement et restent pertinents dans le temps. Cela peut impliquer des coûts supplémentaires pour les correctifs et les ajustements de paramètres. Enfin, la surveillance constante des modèles est également nécessaire pour détecter et corriger les problèmes éventuels, ce qui peut engendrer des coûts supplémentaires en termes de personnel et de temps. En somme, les coûts élevés d'un modèle mal géré peuvent rapidement s'accumuler, entraînant des pertes financières pour l'entreprise.
- **Manque d'explicabilité** : Les modèles mal gérés peuvent également être difficiles à comprendre et à expliquer, ce qui peut rendre difficile pour les utilisateurs finaux et les partenaires de les utiliser et de les faire confiance.

## Les avantages de l'adoption du MLOps pour optimiser la gestion de la chaîne de production de modèles de machine learning

### La définition du MLOps

Inspiré de **DevOps**, Le terme **MLOps** est composé **ML** pour "**Machine Learning**" et **Ops** pour "**Operations**". Il désigne un ensemble de pratiques et de processus qui permettent de gérer efficacement la chaîne de production de modèles de machine learning. Il s'agit d'une approche qui vise à automatiser et à standardiser les différentes étapes de la mise en production de modèles tout en garantissant le bon usage des modèles.



Les pipelines de Machine Learning comprennent des étapes de préparation des données, d'entraînement de modèles, d'évaluation et de déploiement en production.

Le MLOps vise à gérer efficacement ces tâches pour assurer la qualité et la performance des modèles en production.

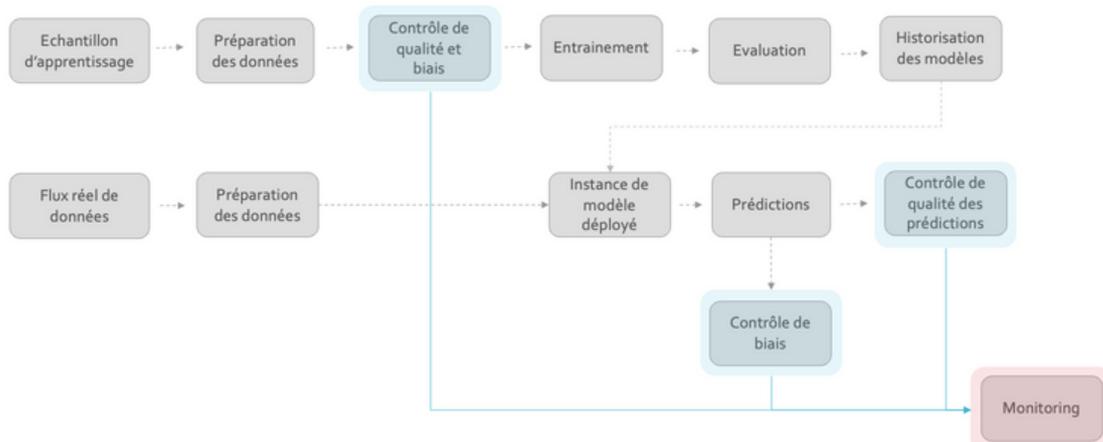


Illustration d'un pipeline de Machine Learning typique

## Les avantages de l'adoption du MLOps pour optimiser la gestion de la chaîne de production de modèles de machine learning

### Comment le MLOps peut aider à résoudre les défis ?

**Automatisation et standardisation :** le MLOps permet d'automatiser et de standardiser les différentes étapes de la chaîne de production de modèles de machine learning, ce qui permet de réduire les coûts et de gagner en temps. Cela inclut :

- La mise en place de pipelines machine learning automatisés pour faciliter le développement, les tests et les déploiements des modèles (CI/CD, IaC, etc.)
- La mise en place d'un "feature store" qui consiste à centraliser le calcul, le stockage et la mise à jour des caractéristiques fréquemment utilisées, rendant ainsi ces dernières disponibles pour une réutilisation par tous les utilisateurs
- La mise en place d'un "components store" qui consiste à centraliser le stockage de bibliothèques de traitement de données spécifiques à l'entreprise, des fonctions de prétraitement de données couramment utilisées pour nettoyer, normaliser ou encoder les données, des fonctions de transformation de données pour créer des features, des modèles pré-entraînés pour des tâches courantes comme la reconnaissance de la parole ou de l'image, ... dans un espace commun
- L'utilisation de technologies de conteneurisation pour faciliter la mise en place de modèles sur des infrastructures distribuées (Vertex.ai, SageMaker, ...)
- L'utilisation de systèmes de gestion de code source pour faciliter la collaboration entre les équipes de développement et de production (Git)
- L'utilisation d'outils pour gérer les versions des modèles et les données utilisées pour les entraîner (model registry)
- La mise en place des processus de contrôle de qualité et de détection de biais sur les données, mais également sur les modèles
- La mise en place de systèmes de monitoring pour surveiller, la qualité et la fiabilité des données, mais aussi les performances des modèles en production.

## Les avantages de l'adoption du MLOps pour optimiser la gestion de la chaîne de production de modèles de machine learning

### Comment le MLOps peut aider à résoudre les défis ?

**Surveillance et maintenance :** le MLOps permet de surveiller et de maintenir efficacement les modèles en production, ce qui permet de garantir la qualité et les performances des modèles, ainsi que de détecter et de résoudre les problèmes éventuels. Cela inclut :

- La mise en place des outils de surveillance qui permettent de surveiller en temps réel les performances des modèles et de détecter rapidement les problèmes potentiels
- La configuration des alertes pour informer immédiatement l'équipe des opérations en cas de problème
- La mise en place des tests automatisés pour vérifier régulièrement que les modèles continuent de fonctionner correctement

**Gouvernance et conformité :** Le MLOps permet de garantir la conformité des modèles aux normes et aux réglementations en vigueur, ce qui permet de réduire les risques juridiques. Cela inclut :

- La vérification de la qualité des données utilisées pour entraîner les modèles, afin de s'assurer qu'elles ne contiennent pas de biais et qu'elles respectent les normes et les usages appropriés.
- La vérification de la conformité des modèles avec les normes en vigueur, comme les normes de qualité pour les applications critiques, pour garantir que les modèles sont sûrs et fiables pour leur utilisation spécifique.
- L'applicabilité des modèles, pour garantir que les décisions prises par les modèles peuvent être comprises et justifiées, ce qui est souvent une exigence réglementaire pour certains domaines d'application.

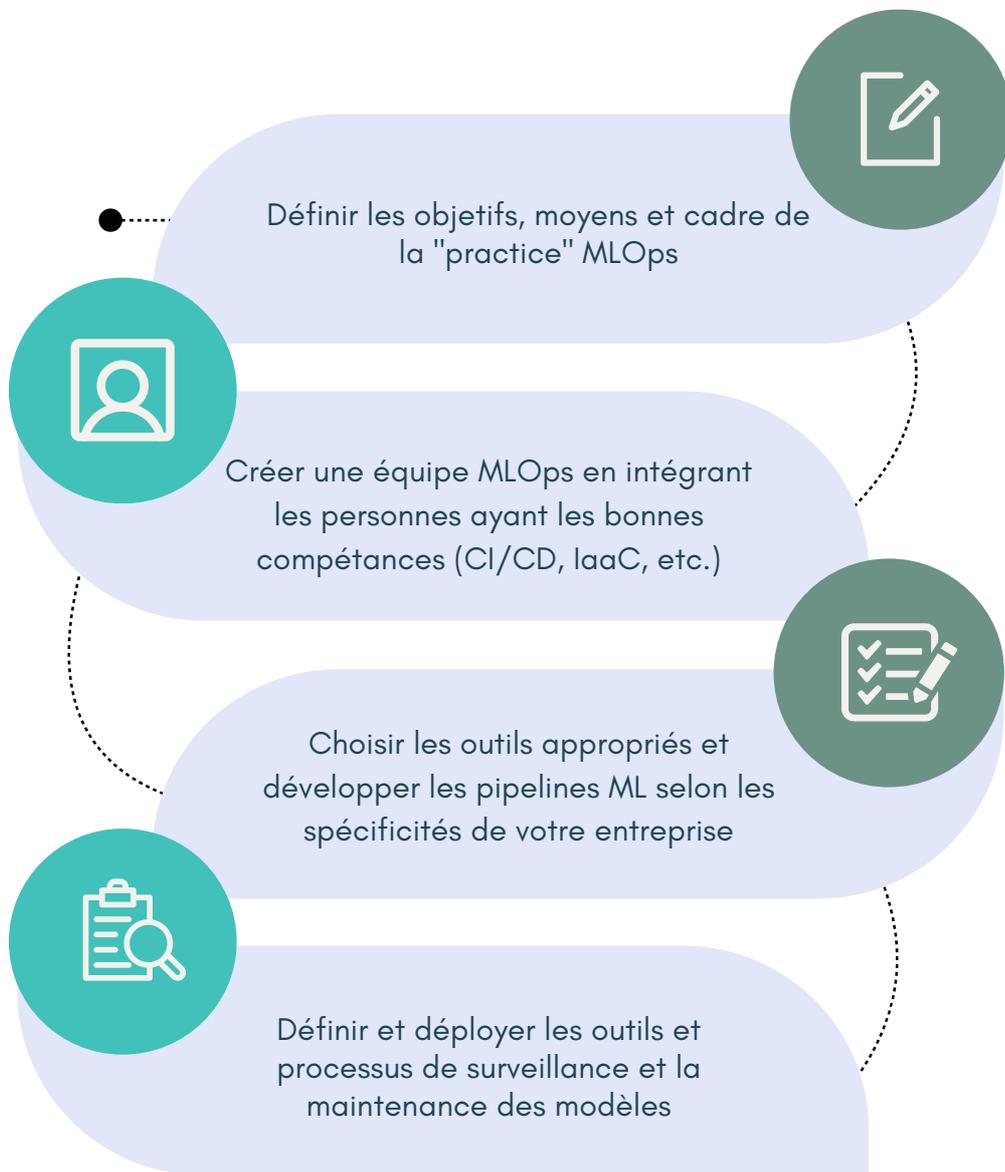
**Collaboration inter-équipe :** Le MLOps permet de faciliter la collaboration entre les différents acteurs impliqués, ce qui permet de garantir que les modèles développés répondent aux besoins et aux objectifs des utilisateurs finaux. Cela inclut :

- La mise en place des outils de collaboration pour faciliter la communication et la coordination entre les équipes de développement, les data scientists et les utilisateurs finaux.
- La mise en place des processus de revue de code garantir la qualité et la fiabilité des modèles en vérifiant leur code et de s'assurer que le code est facile à comprendre et à maintenir.

# 03

## La mise en place une stratégie de MLOps

### Les étapes



## La mise en place une stratégie de MLOps

Les points clés à prendre en compte

### **Automatisation et standardisation**

Automatiser et de standardiser les différentes étapes de la chaîne de production pour garantir la qualité et les performances des modèles.

### **Flexibilité et scalabilité**

Mettre en place, dans la limite du possible, des outils et processus flexibles et évolutifs pour s'adapter aux besoins changeants de l'entreprise.

### **Surveillance et maintenance**

Garantir la capacité de surveiller efficacement la qualité et les performances des modèles en production.

### **Gouvernance et conformité**

Garantir aussi la conformité des modèles aux normes et aux réglementations en vigueur pour réduire les risques juridiques.

### **Collaboration inter-équipe**

Favoriser la collaboration entre les différents acteurs impliqués dans la mise en production de modèles, en utilisant des outils et des processus adaptés.

### **Accessibilité**

Fournir une documentation adéquate pour permettre aux membres de l'équipe de comprendre et de mettre en œuvre les processus et les outils de MLOps.

## La mise en place une stratégie de MLOps

Les 3 erreurs à éviter

### Ne pas impliquer suffisamment les différents acteurs

Il est important d'impliquer les différents acteurs de l'entreprise dans la mise en place de la stratégie de MLOPS pour garantir la réussite de cette dernière. Voici les principaux acteurs :

- **les data scientists**: ils sont responsables de la création et de l'entraînement des modèles,
- **les ingénieurs logiciels**: ils sont responsables du développement des pipelines de production pour les modèles,
- **les responsables de l'exploitation**: ils sont responsables de la mise en place et de la surveillance des modèles en production,
- **les responsables de la gouvernance des données** : ils sont responsables de la gestion des données utilisées par les modèles
- **Les responsables métier** : ils sont responsables de l'utilisation des modèles dans les applications métier

### Ne pas prendre en compte les besoins spécifiques de l'entreprise

Il est important de prendre en compte les besoins spécifiques de l'entreprise pour choisir les outils et les processus appropriés.

### Ne pas avoir de plan de continuité d'activité

Il est crucial de disposer d'un plan de continuité d'activité pour assurer la disponibilité des modèles en cas de panne. Cela inclut la mise en place de procédures de "*disaster recovery plan*", la création de sauvegardes et la mise en place de processus pour tester régulièrement ces plans.

# A propos de l'auteur



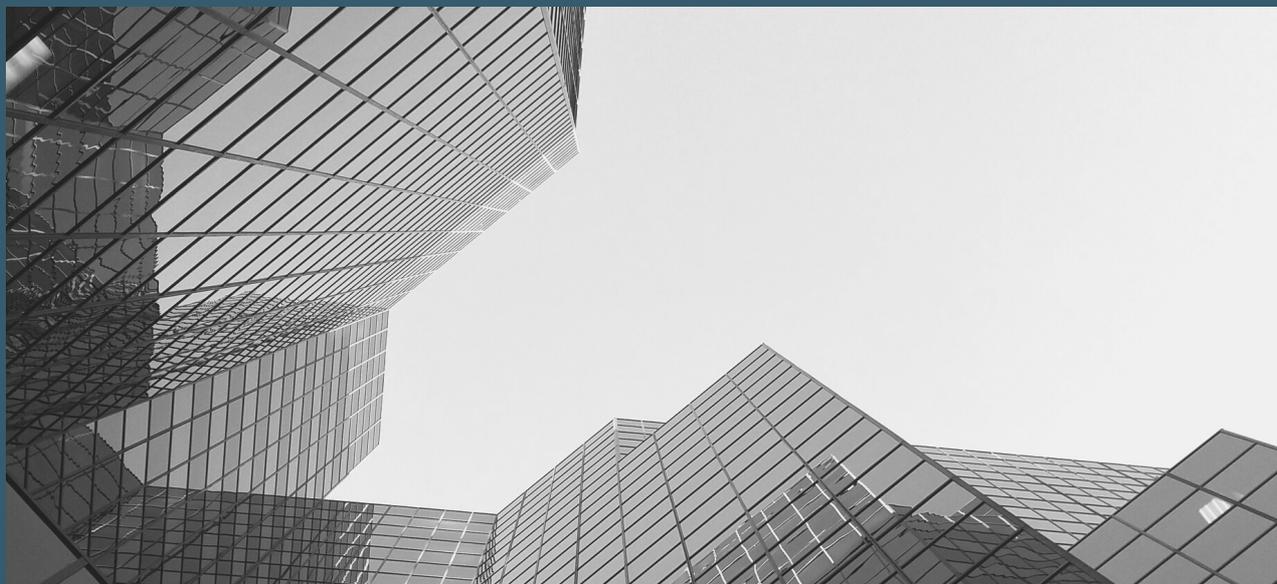
Abdoul-Raoufou GAMBO  
CEO de Kheopsys

Depuis 10 ans, j'interviens sur la définition et la mise en œuvre des stratégies data destinées à améliorer l'efficacité opérationnelle des organisations.

J'ai eu l'opportunité d'accompagner des sociétés telles que Carrefour, Pierre-Fabre, AccorHotels, Hyundai et FM Games à améliorer leurs performances business grâce à l'usage de la data.

**Avec Kheopsys, nous apportons une expertise data, business et technologie aux organisations afin qu'elles puissent tirer le meilleur parti de leur data.**





J'espère que vous avez apprécié la lecture de notre livre blanc et que vous avez trouvé des informations utiles pour améliorer vos processus d'exploitation de la data.

Je tiens à vous remercier pour votre temps et votre intérêt. Si vous avez des commentaires ou des suggestions, n'hésitez pas à me les envoyer par e-mail: [agambo@kheopsys.com](mailto:agambo@kheopsys.com)

Enfin, si vous souhaitez accélérer les décisions data-driven génératrices de performance business, n'hésitez pas à nous contacter. Nous pouvons vous accompagner dans votre projet pour vous aider à atteindre vos objectifs de performance business. N'attendez plus et prenez contact avec nos experts.



54 rue de Londres, 75008 Paris



+337 56 93 39 09



<https://kheopsys-technology.com>

