



L'avantage de l'IA générative :

Guide sur l'utilisation des données comme facteur de différenciation à l'intention des fondateurs



Table des matières

PRÉSENTATION

Guide sur l'utilisation des données comme facteur de différenciation à l'intention des fondateurs	3
---	---

L'IA GÉNÉRATIVE

Une brève introduction	4
Mettre les données au centre de votre approche en matière d'IA générative	5

SECTION 1

Convertir les données en un facteur de différenciation pour les applications d'IA générative	7
--	---

SECTION 2

Établir une fondation de données pour l'IA générative	11
---	----

SECTION 3

Ne pas se limiter à la technologie pour profiter d'un avantage concurrentiel	14
--	----

CONCLUSION

Exploiter les données d'une toute nouvelle façon pour en maximiser la valeur	19
--	----

GLOSSAIRE	20
-----------------	----

PRÉSENTATION

Guide sur l'utilisation des données comme facteur de différenciation à l'intention des fondateurs

L'intelligence artificielle (IA) générative fascine le monde entier grâce à son potentiel de transformer nos modes de communication, de création et de fonctionnement. Ce sont, en grande partie, les applications destinées aux consommateurs, comme ChatGPT et YouChat, qui ont suscité cet attrait pour l'IA générative. Ces robots conversationnels révolutionnaires ont démontré leur capacité impressionnante à imiter la créativité et les conversations des humains.

Mais la véritable force de l'IA générative transcende les robots conversationnels aux connaissances générales. Les jeunes entreprises, tous secteurs confondus, commencent à peine à découvrir les capacités de l'IA générative ainsi que les façons d'innover qui en découlent. Elles sont désireuses de saisir ces possibilités, et pour cause. Selon [Goldman Sachs](#), l'IA générative pourrait faire augmenter le produit intérieur brut (PIB) mondial de près de 7 billions d'USD et hausser la productivité de 1,5 point de pourcentage sur une période de dix ans. En même temps, la technologie attire les investisseurs. En effet, selon [Pitchbook](#), les jeunes entreprises qui se spécialisent dans l'IA générative ont recueilli 27 milliards d'USD en 2023.

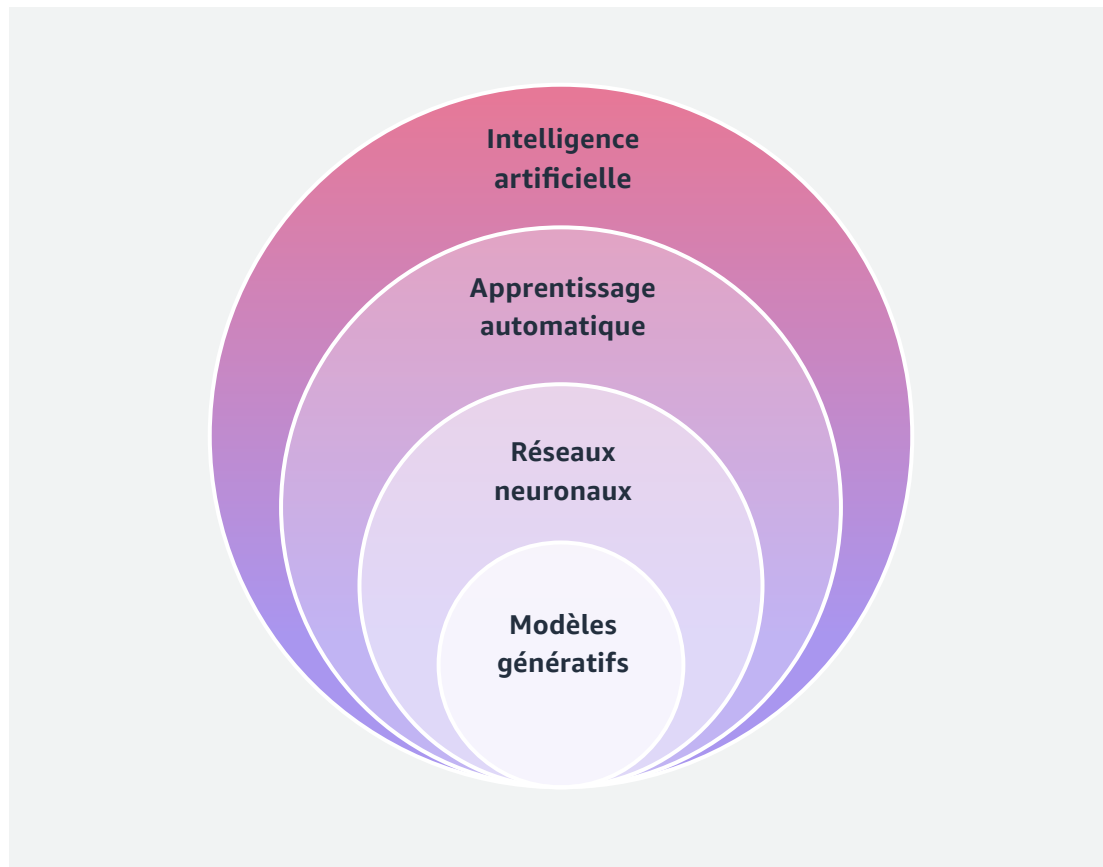
À mesure que la technologie se développe, une nouvelle vague de jeunes entreprises et d'entreprises technologiques novatrices accélère le déploiement de nouvelles capacités d'IA générative en vue de répondre à la demande croissante des clients et des entreprises. Cette innovation continue entraîne une augmentation des cas d'utilisation d'IA générative disponibles.

Devant de tels résultats, il n'est donc pas étonnant que les fondateurs de jeunes entreprises et les responsables des données souhaitent mettre rapidement en place leurs propres applications d'IA générative. Ils veulent non seulement être en mesure de passer à la prochaine étape, mais également d'obtenir un avantage concurrentiel dans ce domaine émergent et d'attirer des investisseurs. Les données de votre jeune entreprise sont la clé pour libérer le potentiel de l'IA générative.

En matière d'IA générative, vos données sont votre facteur de différenciation.

Une brève introduction

De façon générale, on peut définir l'IA générative comme un type d'IA utilisé pour favoriser l'émergence d'idées nouvelles et la création de contenus inédits. Les applications d'IA générative peuvent, par exemple, rédiger des articles, générer du code et concevoir des images numériques. Elles permettent également d'automatiser des tâches laborieuses, comme la condensation de longs documents en résumés.



La liste est loin d'être exhaustive. Cependant, les clients utilisent fréquemment des applications d'IA générative pour accomplir les tâches qui suivent :

- Améliorer l'expérience client en mettant à profit des capacités comme des robots conversationnels, des assistants virtuels, la personnalisation ou la modération de contenu.
- Augmenter la productivité des employés grâce à la recherche conversationnelle, à la création de contenu, à la synthèse de textes ou à la création de code.
- Mettre le turbo dans la production de contenu créatif de tout type, y compris les œuvres d'art, la musique ou les animations.
- Simplifier les opérations commerciales grâce au traitement intelligent de documents, à la maintenance prédictive, au contrôle de la qualité et à l'inspection visuelle ou à l'augmentation de données.

L'IA GÉNÉRATIVE

Comme toutes les IA, l'IA générative est alimentée par des modèles d'apprentissage automatique (AA) : de très grands modèles préalablement entraînés sur d'énormes quantités de données. Ces modèles sont couramment appelés [modèles de fondation \(FM\)](#).

Il importe de souligner que, à la base, un FM met à profit les dernières avancées en matière d'AA. Il existe une catégorie de FM, comme les modèles de transformateurs génératifs préalablement entraînés (GPT), que l'on nomme communément grands modèles de langage (LLM). Ces modèles se concentrent spécialement sur des tâches linguistiques, comme la synthèse, la génération de texte et les questions et réponses ouvertes. Les LLM sont particuliers parce qu'ils contiennent un grand nombre de paramètres qui leur permettent d'apprendre des concepts avancés.

Mettre les données au centre de votre approche en matière d'IA générative

Nous savons depuis longtemps que les données sont un atout stratégique pour les entreprises. Pourtant, selon une étude d'[Accenture](#), seule une entreprise sur trois, soit 32 %, se dit en mesure de dégager une valeur tangible et mesurable des données, même après avoir investi dans l'infrastructure de données.

L'IA générative peut donner une impulsion à ces chiffres. Elle vous offre une occasion d'exploiter vos données d'une nouvelle façon et d'en maximiser la valeur. La technologie vous permet d'accélérer l'innovation à l'aide de vos données, d'utiliser ces données dans de nouveaux types d'applications et d'exploiter la valeur des données qui ont toujours été difficiles à utiliser, comme les données non structurées.

Certains de nos clients ont déjà commencé à combiner les données à l'IA générative pour améliorer leurs résultats commerciaux et l'expérience client. Ainsi, Intuit a conçu [Intuit Assist](#), un nouvel assistant génératif alimenté par l'IA générative qui utilise des jeux de données contextuelles pertinentes touchant les petites entreprises, le crédit à la consommation et la fiscalité dans le but de fournir des informations financières personnalisées aux clients.

Jeune entreprise offrant un programme d'aide aux employés (PAE) novateur, [Clearhead](#) a voulu accroître l'accès à du soutien en santé mentale en facilitant la recherche du bon thérapeute pour les utilisateurs de sa plateforme de bien-être. Pour y arriver, elle a notamment amélioré la réactivité de son robot conversationnel de soutien psychologique numérique. Clearhead a utilisé [Amazon Titan](#) sur [Amazon Bedrock](#) pour améliorer la personnalisation du robot conversationnel, ce qui a mené à des discussions plus sensibles au contexte, à une plus grande précision dans les recherches de thérapeutes, à une réduction des coûts d'IA générative et à des réponses plus prévisibles.

L'IA GÉNÉRATIVE

Ces applications sont porteuses et ne représentent sans doute qu'une infime fraction de ce que l'IA générative peut offrir aux jeunes entreprises et à leurs clients. Les transformations que l'IA générative réserve à notre monde sont toujours en train de se préciser. Peut-être peinez-vous à comprendre comment saisir cet énorme potentiel, alors même que nous avons l'impression de nous aventurer en terrain inconnu. En réalité, la voie à suivre pour dégager une valeur commerciale de l'IA générative ne diffère guère de celle que nous prenons avec les autres technologies. Tout repose sur votre stratégie de données et sur la façon dont vous vous y prenez pour faire de vos données un facteur de différenciation au sein de cette stratégie.

Il se peut que vous ayez déjà une stratégie de données en place ou que vous commenciez tout juste à en élaborer une. Dans les deux cas, le moment est tout indiqué pour rassembler cette stratégie de manière à vous permettre de propulser votre valeur commerciale grâce à l'IA générative. C'est une occasion sans précédent qui s'offre à vous d'obtenir un avantage concurrentiel durable en vous servant de vos données pour vous démarquer.

Dans ce livre blanc, les fondateurs et leurs équipes trouveront des informations et les prochaines étapes à suivre pour mettre leurs données au service de la création d'applications d'IA générative qui sont uniques à leur jeune entreprise. Pour innover et être concurrentiel dans ce domaine, vous devez élaborer une vaste stratégie de données, qui englobe la technologie en plus de vos priorités opérationnelles, de vos cas d'utilisation, de vos employés et de vos barrières de protection en matière de gouvernance. Considérée dans son ensemble, cette stratégie se veut une vision moderne des données qui vous permet de dégager une valeur commerciale de vos applications d'IA générative.

Pour vous aider dans la création de cette stratégie de données moderne, nous miserons sur trois domaines :

1

Convertir vos données en un facteur de différenciation pour vos applications d'IA générative

2

Établir la bonne fondation de données pour exploiter la valeur de vos données actuelles grâce à l'IA générative

3

Ne pas se limiter à la technologie pour profiter d'un avantage concurrentiel grâce à l'IA générative

SECTION 1

Convertir les données en un facteur de différenciation pour les applications d'IA générative

Vos données font la différence entre une application d'IA générative générique et une application qui connaît profondément votre jeune entreprise et vos clients. C'est pourquoi vous devrez déterminer la meilleure façon d'utiliser vos données pour bien rendre compte du caractère unique de votre jeune entreprise et le mettre en valeur. Pour la plupart des jeunes entreprises, le déploiement d'applications d'IA générative commence par un modèle de fondation (FM) prêt à l'emploi. Une poignée de jeunes entreprises choisissent de créer leurs propres FM pour alimenter leurs applications d'IA générative. Pour ce faire, elles ont toutefois besoin de vastes ressources informatiques et d'un personnel hautement spécialisé.

Malgré leur puissance, les FM prêts à l'emploi sont généralisés de par leur conception. Comme leur nom l'indique, ce sont des fondations. Ainsi, ils ne sont pas adaptés aux besoins de votre jeune entreprise, car ils ne peuvent accéder à ses données les plus récentes ni effectuer des tâches propres à un domaine pour répondre aux demandes des utilisateurs. Vos données sont essentielles pour harmoniser les applications d'IA générative avec votre expérience client, vos connaissances internes, la voix de votre marque et vos paramètres éthiques.

Par exemple, si vous êtes une agence de voyages en ligne qui souhaite se servir d'une application d'IA générative pour améliorer les recommandations qu'elle donne à ses clients, vous aurez probablement tout intérêt à utiliser des données propres à vos clients individuels, comme leurs voyages antérieurs, leur historique de consultation Web et leurs préférences en matière de voyage. Vous souhaitez également accéder à des données agrégées sur des profils de voyageurs et des inventaires de voyages similaires afin d'améliorer la recommandation. En exploitant vos données, vous créez une expérience client personnalisée et unique.

De plus, les FM prêts à l'emploi sont faciles à se procurer. Vos applications d'IA générative peuvent ainsi se démarquer lorsque vous les personnalisez au moyen de vos propres données. Supposons que vous utilisez également un FM prêt à l'emploi pour rédiger un article de marketing pour votre agence de voyages en ligne. Il se peut que vos concurrents utilisent le même modèle pour obtenir le même effet que vous. Ces modèles s'inspirent en grande partie du même bassin de connaissances générales. Sans personnalisation, vous pourriez finir par créer du contenu qui ressemblerait étrangement à celui de vos concurrents et vice-versa.

La personnalisation crée un avantage concurrentiel durable. Quelques méthodes peuvent vous aider à personnaliser votre FM, y compris le réglage fin et l'apprentissage en contexte.

SECTION 1

Pleins feux sur un client

NRIX, un fournisseur mondial de données et d'analyses de transport, met au point une nouvelle solution centrée sur [Amazon Bedrock](#). Cette solution fournira des informations à jour et en temps réel pour aider les ingénieurs en circulation et en sécurité à comprendre où, quand et pourquoi un événement survient dans une rue et la manière d'y réagir.

Sa nouvelle solution Amazon Bedrock utilise la génération augmentée de récupération (RAG) pour augmenter les requêtes au FM sous-jacent au moyen de données historiques, comme l'historique des incidents et des accidents liés aux excès de vitesse, et des données récentes comme l'état des embouteillages et les conditions météorologiques du moment.

En utilisant ses propres données pour améliorer son FM dans Amazon Bedrock, INRIX peut fournir à ses clients des réponses rapides à des questions complexes, comme les changements à apporter aux routes pour réduire les embouteillages et diminuer le nombre d'accidents, la façon de déterminer l'emplacement idéal d'un nouveau magasin de détail ou même la façon d'atténuer les problèmes de circulation et de stationnement lors d'un concert à venir.



SECTION 1

Modèles de données d'IA générative émergents



Réglage fin

Avec un FM prêt à l'emploi, vous devez utiliser vos données pour personnaliser le modèle en fonction de vos besoins commerciaux uniques. Le réglage fin est une bonne option pour les applications qui demandent une connaissance élevée du domaine, telles que les agents de soutien technique ou la création de contenu propre à votre entreprise. Grâce à Amazon Bedrock, vous pouvez personnaliser en toute sécurité un FM à partir de vos données et utiliser d'autres outils intégrés pour créer des applications qui connaissent votre entreprise, vos données et vos clients.

Imaginez un directeur du marketing de contenu qui travaille dans une jeune entreprise axée sur le commerce électronique. Ce dernier doit concevoir des publicités et des campagnes originales et ciblées pour une nouvelle gamme de sacs à main à sortir prochainement. Pour ce faire, il fournit à Amazon Bedrock quelques exemples étiquetés qu'il a tirés des slogans les plus percutants des campagnes précédentes ainsi que les descriptions des produits qui y sont associées, le tout hébergé dans un lac de données sur [Amazon Simple Storage Service \(Amazon S3\)](#). Amazon Bedrock fait une copie distincte du modèle de base à laquelle seul le client peut accéder à des fins d'entraînement du modèle. À l'issue de l'entraînement, Amazon Bedrock génère automatiquement du contenu efficace sur les nouveaux sacs à main à publier sur les médias sociaux, les placards publicitaires et le Web.

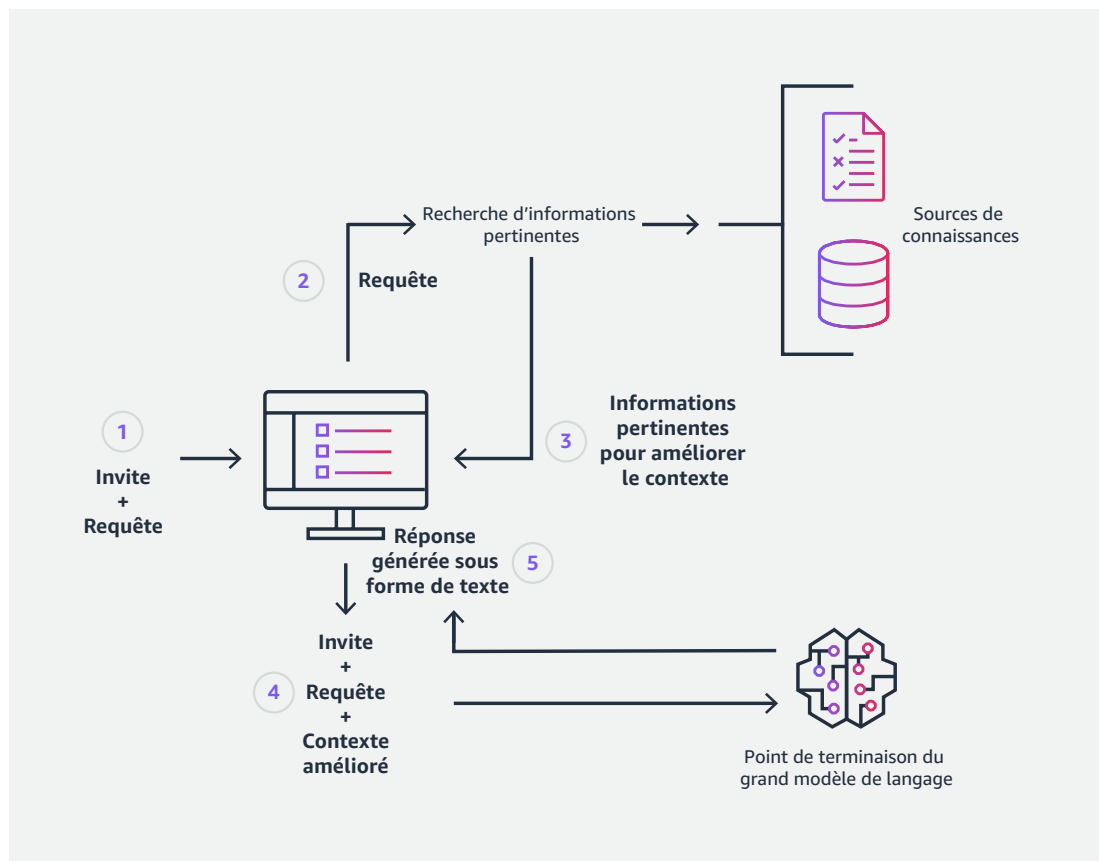
SECTION 1

Apprentissage en contexte

Les modèles de fondation sont entraînés à un moment précis, et il n'est pas utile de procéder à un réglage fin à chaque changement d'un jeu de données. Une fois qu'ils sont entraînés, ils n'ingèrent plus de nouvelles connaissances ni de nouvelles données. S'ils ont besoin d'un contexte plus précis pour résoudre un problème, ils n'arrivent pas à trouver de l'information en temps réel ni à y accéder.

Pour accroître la pertinence et la contextualisation des réponses, vous pouvez former votre FM avec des données grâce à l'apprentissage en contexte, une technique dans laquelle le FM est guidé vers des données contextuelles propres à un domaine, soit par l'ingénierie de requête, soit par la RAG. De nombreuses entreprises feront de la RAG la principale méthode d'apprentissage en contexte. La RAG permet à votre FM d'accéder aux données les plus récentes de votre jeune entreprise pour donner une réponse plus précise et plus pertinente. La RAG utilise bien souvent des intégrations vectorielles ou des représentations numériques de mots, de phrases ou d'images. Les intégrations codent la signification sémantique du texte source ou des images, ce qui permet aux FM de trouver plus facilement des liens entre des vecteurs similaires et d'améliorer les réponses aux requêtes.

Ces techniques peuvent certes être utilisées séparément; cependant, la combinaison du réglage fin et de la RAG peut vous aider à convertir vos données en un facteur de différenciation pour vos applications d'IA générative.



SECTION 2

Établir une fondation de données pour l'IA générative

Vos données sont primordiales pour créer de la valeur avec les applications d'IA générative. Il devient alors vital de personnaliser vos modèles au moyen de données de haute qualité, pertinentes, facilement accessibles et disponibles. Pour satisfaire ces critères, vous devez d'abord disposer d'une fondation de données solide, qui comprend un ensemble exhaustif et intégré de services de données pour répondre à toutes les applications et tous les cas d'utilisation et types de données en plus des outils pour gouverner ces données. Voici un aperçu général de cette fondation de données.

Complète

Pour l'IA générative, vous devez stocker différents types de données, y compris des données non structurées, structurées, de diffusion en continu et vectorielles, qui peuvent être utilisées pour créer et personnaliser des modèles et ajouter du contexte aux requêtes, avec ou sans RAG. Un ensemble complet de services de données permet de stocker toutes ces données ainsi que de les interroger et de les analyser à l'échelle.

En règle générale, un ensemble complet de services de données pour l'IA générative comprend un lac de données hautement durable et évolutif. Ce lac de données stocke les données propres au domaine dont vous avez besoin pour créer et personnaliser vos FM. Depuis des années, Amazon Web Services (AWS) aide ses clients à jeter des bases solides aux lacs de données afin d'y stocker des données structurées et non structurées grâce à des services comme [Amazon S3](#), [AWS Glue](#) et [AWS Lake Formation](#). Les clients utilisent Amazon S3 pour créer des centaines de milliers de lacs de données.

De même, une fondation de données pour l'IA générative comprend des bases de connaissances à haute performance pour la RAG. AWS vous propose plusieurs options en fonction de votre cas d'utilisation. Par exemple, les bases de données NoSQL stockent l'état et l'historique des conversations. Un robot conversationnel peut donc se souvenir des réponses antérieures. Les bases de données transactionnelles stockent le contexte et les informations sur les clients afin de personnaliser davantage les réponses. Vous pouvez également utiliser une base de connaissances comme [Amazon Kendra](#) qui se connecte à plusieurs référentiels de contenu structuré et non structuré. Vos FM disposent alors d'une source de connaissances qui repose sur l'accès à des documents. Vous pouvez également recourir à des bases de données dotées de capacités de recherche vectorielle, spécialement conçues pour stocker et récupérer efficacement des intégrations. La possibilité d'utiliser la recherche vectorielle dans les bases de données que vous utilisez déjà comporte plusieurs avantages. Par exemple, elle élimine la courbe d'apprentissage abrupte pour les nouveaux outils de programmation, interfaces de programmation d'application et kits de développement logiciel. En outre, vous savez que vos bases de données ont déjà fait leurs preuves en production et répondent aux exigences en matière de capacité de mise à l'échelle, de disponibilité, de stockage et de calcul, ce qui vous met en confiance. De même, lorsque vos vecteurs et vos données d'entreprise sont stockés au même endroit, vos applications peuvent s'exécuter plus rapidement. Vous n'avez pas à vous soucier de la synchronisation des données ni de leur migration. AWS offre des fonctionnalités de recherche vectorielle pour bon nombre de ses magasins de données les plus utilisés afin d'offrir aux clients encore plus de flexibilité dans la création de leurs applications d'IA générative.

SECTION 2

Une fondation de données complète tient compte à la fois de l'analytique et du stockage des données. Vous pouvez utiliser des entrepôts de données pour personnaliser vos FM et les cas d'utilisation qui exigent des données opérationnelles à jour, comme la création d'un FM ou d'un autre LLM pour fournir des informations sur les données commerciales au moyen de requêtes en langage naturel. [Amazon Redshift](#) est un entrepôt de données rapide à l'échelle du pétaoctet, qui offre un rapport prix-performances jusqu'à six fois supérieur à celui des autres entrepôts de données dans le nuage¹ Amazon Redshift s'intègre à bon nombre de vos sources de données, y compris Amazon S3 et [Amazon Aurora](#), ce qui vous donne un aperçu plus complet de vos données.

« Les entreprises qui n'ont pas encore trouvé de moyens d'harmoniser efficacement leurs données et de fournir un accès facile à leurs données seront incapables d'affiner l'IA générative pour exploiter davantage ses utilisations potentiellement transformatrices. »

What every CEO should know about generative AI, McKinsey Digital, 2023

Intégrée

L'intégration des données vous donne un aperçu complet de votre entreprise et garantit l'accès facile à vos données pour vos applications d'IA générative. Grâce aux intégrations directes entre les services AWS, nous réduisons et éliminons l'extraction, la transformation et le chargement (ETL) pour les cas d'utilisation courants, ce qui permet à votre équipe d'accélérer la cadence. Vous pouvez également utiliser AWS Glue, notre service d'ETL et d'intégration de données sans serveur et évolutif qui facilite la découverte, la préparation, le déplacement et l'intégration de données provenant de plusieurs sources pour l'analytique et l'apprentissage automatique. AWS se connecte à des centaines de sources de données, y compris des logiciels en tant que service, sur site et à d'autres nuages, ainsi qu'à des données tierces provenant de plus de 300 fournisseurs de données.

SECTION 2

Sécurisée et gouvernée

Vos données doivent être sécurisées et gouvernées tout au long du cycle de vie de la création d'une application d'IA générative. AWS met à votre disposition plusieurs outils qui vous aident à garantir la qualité des données, la confidentialité et les contrôles d'accès. Les données que vous utilisez pour les applications d'IA générative sont donc conformes et de haute qualité.

Avec Amazon Bedrock, toutes vos données peuvent être chiffrées au repos à l'aide de vos propres clés Key Management Service (KMS), ce qui offre une visibilité et un contrôle complets sur le stockage de vos données et modèles personnalisés et sur l'accès à ces derniers. Grâce à [AWS PrivateLink](#), vous pouvez transmettre vos données AWS à Amazon Bedrock exclusivement par le réseau AWS et jamais par l'Internet public. Vous pouvez personnaliser en privé un FM sur votre propre nuage privé virtuel (VPC). Ainsi, il n'y a jamais de fuite de données et aucune donnée n'est utilisée pour entraîner ou personnaliser un modèle qui serait à la disposition d'autres jeunes entreprises. Les FM [Amazon Titan](#) sont conçus pour détecter et supprimer les contenus préjudiciables dans les données que vous voulez utiliser à des fins de personnalisation, rejeter les contenus inappropriés dans les entrées de l'utilisateur et filtrer les sorties du modèle qui contiennent du contenu inapproprié, comme des injures ou des discours haineux.

Précédemment, nous avons abordé l'importance de personnaliser votre FM pour qu'il soit à l'image de votre marque, qu'il traduise votre expérience client et qu'il élimine le contenu inexact ou non pertinent. La personnalisation est certes importante pour relever ces défis. Cependant, la gouvernance des données sous forme de surveillance humaine et de rétroaction l'est tout autant. Pour garantir que votre résultat reflète la façon dont vous souhaitez être perçu, vous aurez besoin de collaborateurs humains. Par exemple, la méthode d'apprentissage par renforcement à partir de la rétroaction humaine vous permet d'entraîner un FM à prendre des décisions et à agir tout en recevant des conseils d'experts humains. Ces experts sont à l'affût de complications potentielles, telles que des biais de données, des problèmes de qualité des données et des lacunes dans les données, et favorisent l'harmonisation de votre FM avec la voix de votre marque, les lignes directrices, l'éthique et les politiques de votre entreprise.

SECTION 3

Ne pas se limiter à la technologie pour profiter d'un avantage concurrentiel

La technologie est une grande facilitatrice pour l'IA générative. Toutefois, elle ne représente qu'un seul élément d'une stratégie visant à faire des données un atout stratégique de votre jeune entreprise.

Chez AWS, nous adoptons une vision moderne, de nature plus vaste, à l'égard de la stratégie de données. Une stratégie de données de bout en bout englobe la technologie en plus de la mentalité, des personnes et des processus. Cette combinaison ouvre la voie à ce que votre équipe devienne axée sur les données par l'intégration de données dans tous les aspects de votre entreprise et de vos activités.

Mentalité

La mentalité s'entend de la façon dont une entreprise perçoit les données et du traitement qu'elle en fait. Elle transparait dans les convictions, valeurs et comportements qui forment sa culture axée sur les données dont les cas d'utilisation se situent dans la même logique.

Traditionnellement, lors de la conception de leurs stratégies de données, les fondateurs voyaient les données comme une plateforme et un moyen de créer des solutions. Cette perspective se traduit systématiquement par une inadéquation entre les investissements en TI et l'amélioration des résultats commerciaux.

Dans un contexte où l'intérêt pour l'IA générative s'accroît, cette façon de penser devient de plus en plus fréquente. Cette tendance est tout à fait normale, car les fondateurs veulent utiliser l'IA générative tout en veillant à demeurer concurrentiels. Ils doivent toutefois commencer par explorer la façon dont ces applications peuvent aider leur entreprise à résoudre un problème ou à se démarquer. Pour y parvenir, ils doivent aborder l'IA générative comme un produit de données évolutif qui leur procure une valeur réelle. Ce faisant, les jeunes entreprises peuvent mettre l'accent sur les clients, et non sur les solutions, et combler l'écart critique entre les initiatives de données et les résultats commerciaux.

« Une jeune entreprise qui exploite les données comme une ressource servant à stimuler l'innovation continue et à créer des connaissances réalisables dynamisant l'expérience des consommateurs, qui en redemanderont plus. »

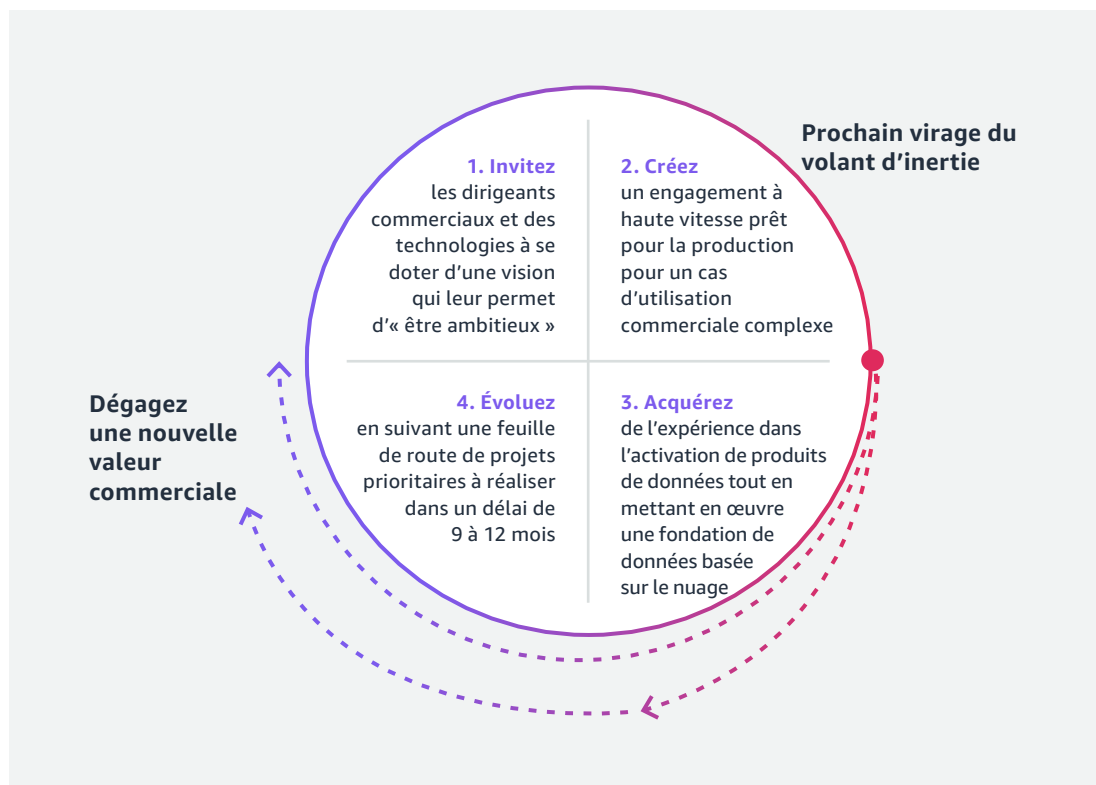
Accelerating business value with a modern data strategy, AWS, 2023

SECTION 3

L'approche du volant d'inertie

Nous aidons nos clients à concevoir des produits axés sur les données pour leurs cas d'utilisation en adoptant [l'approche du volant d'inertie](#). Cette approche vous permet d'accélérer le rythme auquel vous détectez une occasion d'affaires ou un nouveau client à forte valeur ajoutée et auquel vous créez un produit ou une expérience de données pour saisir cette occasion. Une fois que le volant d'inertie se met à tourner, vous continuez à trouver de nouveaux cas d'utilisation pour accroître le nombre de produits ou d'expériences que vous offrez.

Le volant accélère la transformation et crée une valeur croissante. En résumé, cela revient à « être ambitieux, commencer modestement et évoluer rapidement ». Nous l'avons illustré brièvement dans l'exemple qui suit.



Par exemple, le détaillant de produits de beauté en ligne [BEAUTY BAY](#) entend offrir une expérience client hors pair et tenir son jeune public informé des dernières tendances. En collaboration avec [BJSS](#), [un Partenaire AWS](#), il a procédé à une transformation numérique qui a visé de nombreux aspects de ses activités. De concert, ils ont mené un examen [AWS Well-Architected](#) du parc informatique, remanié le système de gestion des commandes et créé une plateforme de données native du nuage sur AWS.

Pour le développement de la plateforme de données, BJSS a utilisé une approche tout en tant que code et l'[AWS CDK](#), un cadre de développement logiciel à code source ouvert qui définit des ressources infonuagiques à l'aide de langages de programmation familiers. La plateforme est désormais plus flexible, sa maintenance est moins chère et plus facile, et son déploiement est plus rapide, ce qui a permis d'améliorer la capacité à innover de l'équipe informatique.

SECTION 3

Donnez le moyen d'agir à vos employés

L'IA générative, tout comme l'IA en général, peut aider votre équipe à exploiter les données plus rapidement et de manière plus ciblée. Au bout du compte, ce changement peut permettre à votre jeune entreprise d'accroître son efficacité et sa productivité en plus d'améliorer l'expérience des employés et des clients.

Par exemple, [Marketing Evolution](#) a créé une solution novatrice de mesure et d'attribution sur AWS. Pour réduire les processus manuels chronophages, Marketing Evolution a commencé à utiliser [AWS Glue](#), un service d'intégration de données sans serveur qui facilite la découverte, la préparation, la migration et l'intégration de données provenant de plusieurs sources pour l'analytique, l'AA et le développement d'applications. Marketing Evolution a réduit ses coûts et augmenté l'efficacité de sa solution, ce qui lui a permis d'accélérer les résultats et d'augmenter le rendement sur capital investi de ses clients.

La jeune entreprise [Protium](#) s'est servie d'[Amazon Aurora](#) pour développer Turiya, sa pile interne de gestion des prêts et des risques centrale à ses activités. Grâce à sa plateforme alimentée par AWS, Turiya offre à Protium la possibilité de personnaliser ses offres de crédit et de les offrir à un plus grand nombre d'utilisateurs. En utilisant un modèle basé sur le flux de travail, l'équipe peut accélérer le lancement de produits et les services sur sa plateforme. De plus, Protium peut désormais se mettre à l'échelle sur tous les canaux et s'y intégrer, le tout en préservant la haute disponibilité, la performance et la conformité de manière rentable.

Rôles émergents dans les domaines de l'IA et de l'AA



Ingénieur en IA générative : se spécialise dans le développement et la mise en œuvre de modèles et de systèmes d'IA générative, y compris la conception, l'entraînement, le réglage fin et l'optimisation de modèles pour des applications données



Curateur de données : assume la responsabilité de l'approvisionnement, de la collecte et de l'organisation de jeux de données de haute qualité qui sont diversifiés, représentatifs et bien étiquetés



Consultant en IA générative : donne des conseils et des directives d'expert sur l'utilisation des technologies d'IA générative dans divers secteurs



Artiste en IA générative : explore les possibilités créatives que recèlent les technologies d'IA générative en vue de produire des œuvres d'art, des compositions musicales ou des conceptions visuelles uniques et novatrices



Spécialiste des politiques et de la réglementation de l'IA : façonne les politiques, les lignes directrices et les règlements qui régissent le développement, le déploiement et l'utilisation responsables des systèmes d'IA générative

SECTION 3

L'IA générative peut aider un plus grand nombre d'employés à mettre les données au profit de l'innovation. Pour éviter les obstacles et les goulots d'étranglement, vous devez vous doter de politiques appropriées qui facilitent la découverte, la consommation, le partage et la gestion des données. Traditionnellement, le service des TI exerçait un contrôle presque absolu sur la diffusion des données au sein d'une jeune entreprise. Une stratégie de données de bout en bout repousse cette responsabilité aux périphéries et la confie aux équipes qui produisent et consomment les données. Nous désignons souvent ce modèle sous le nom de [communauté de données moderne](#) et nous voyons des clients l'utiliser pour permettre à un plus grand nombre d'employés de prendre des décisions fondées sur les données qui touchent leur fonction commerciale particulière.

Nouvelles compétences et nouveaux rôles

L'IA générative ouvrira la porte à la création de rôles entièrement nouveaux et augmentera la demande pour les rôles existants. Selon le [Rapport sur l'avenir de l'emploi 2023 du Forum économique mondial](#), la majorité des postes qui enregistrent la croissance la plus rapide sont axés sur la technologie et comprennent des spécialistes de l'IA et de l'AA en plus d'analystes de données, de scientifiques des données et d'analystes de la sécurité de l'information.

Cependant, à l'heure actuelle, le vivier de talents qui se spécialisent dans l'IA, l'AA et l'infonuagique est limité. Les jeunes entreprises sont souvent privées d'une main-d'œuvre qualifiée et diversifiée, qui leur permettrait d'adopter pleinement leurs stratégies de données. Pour poursuivre sur votre lancée, vous devez investir dans l'amélioration des compétences de vos employés actuels, y compris ceux qui ne relèvent pas du service des TI et qui travaillent au sein d'équipes telles que les finances ou le marketing. En formant ces équipes à interagir avec les données, vous éliminez les goulots d'étranglement, car vous leur donnez accès aux bons produits de données au bon moment pour qu'ils puissent prendre des décisions commerciales.

L'amélioration des compétences des employés est un élément important de votre stratégie de données. Ces exercices sont un investissement, mais ils n'ont pas le poids que certains fondateurs pourraient imaginer. Encore une fois, grâce à l'IA et à l'AA, vous pouvez plus facilement équiper les employés ayant des niveaux de compétence technique divers des outils dont ils ont besoin pour analyser des données, découvrir des informations et concevoir des récits. Par exemple, [Amazon Q](#), notre nouvel assistant d'IA générative, vous aide dans [Amazon QuickSight](#) à créer des tableaux de bord et des histoires visuelles convaincantes à partir des données de votre tableau de bord en langage naturel. Nous avons également annoncé qu'Amazon Q pouvait vous aider à créer des pipelines d'intégration de données en langage naturel. Par exemple, vous pouvez demander à Q de « lire les fichiers JSON à partir de S3, de les joindre sur "accountid" et de les charger dans DynamoDB ». Amazon Q renvoie une tâche d'intégration de données de bout en bout pour accomplir cette action. Avec Amazon Q, les analystes des données, les scientifiques des données et les ingénieurs peuvent également accroître leur productivité en utilisant la fonctionnalité texte à SQL d'IA générative dans Amazon Redshift pour interroger les données contenues dans votre entrepôt de données.

SECTION 3

Créer de nouveaux processus ou optimiser les processus existants pour suivre le rythme de l'innovation

L'IA générative a mis l'expression « [IA responsable](#) » à l'avant-plan et a ajouté une nouvelle dimension à la gouvernance des données. En plus des questions liées au partage des données, à la sécurité et à la protection de la vie privée, vous devez tenir compte de celles liées aux préjugés, à l'éthique et aux fausses informations.

Dans la section précédente, nous avons traité de certains des outils qui permettent d'atténuer les risques communs associés à l'IA générative. Toutefois, comme le savent les fondateurs de jeunes entreprises, la gouvernance des données est davantage une question de stratégie que de sélection d'outils. La gouvernance des données doit être structurée de manière à soutenir une stratégie plus large en matière de données et d'IA, mais mise en œuvre progressivement en fonction des cas d'utilisation et des priorités commerciales.

Les fondateurs nous disent souvent qu'ils ont de la difficulté à démontrer la valeur commerciale de leurs initiatives en matière de gouvernance des données. Cependant, lorsqu'ils examinent de manière plus générale leur stratégie de données et l'harmonisent avec les priorités commerciales, ils se rapprochent de la démonstration de cette valeur.

Une stratégie de données générale est importante, car elle permet de veiller à ce que les politiques de gouvernance des données soient en phase avec les structures organisationnelles. Pour innover et évoluer rapidement grâce à l'IA générative, votre équipe a besoin d'un accès facile aux données. Elle doit également agir dans les limites des politiques de sécurité, de partage et de protection de la vie privée. Vous devez trouver l'équilibre entre les essentiels et l'autonomie et la rapidité. Vous atteignez cet équilibre lorsque vous voyez la gouvernance des données comme une facilitatrice de votre stratégie de données plus générale.



CONCLUSION

Exploiter les données d'une toute nouvelle façon pour en maximiser la valeur

Nous commençons à peine à entrevoir les possibilités de l'IA générative, mais nous percevons déjà qu'elle transformera presque toutes les applications, tous les secteurs et toutes les jeunes entreprises. L'IA générative offre des façons nouvelles et prometteuses d'exploiter vos données existantes et d'en maximiser la valeur. Elle peut vous permettre de mettre au point de nouveaux produits et expériences de données qui raviront vos clients et permettront à votre marque de se démarquer.

Nous savons que presque tous les fondateurs souhaitent explorer le potentiel de l'IA générative et que vos prochaines étapes seront cruciales. Plutôt que d'utiliser une application générique, vous devriez vous concentrer sur ce qui fait l'unicité de votre entreprise et vous en servir pour orienter votre prise de décision. Cette unicité existe au sein de vos données. Pour exploiter pleinement vos données en tant qu'atout stratégique, vous devez créer une stratégie de données qui comprend non seulement votre technologie, mais aussi votre mentalité, vos employés et vos processus. Chaque couche de cette stratégie vient consolider fortement vos données en tant que différenciateur de la façon dont vous créez et personnalisez vos applications d'IA générative et de la façon dont vous donnez à votre équipe les moyens d'innover.

Apprenez-en plus sur la création d'une fondation de données pour obtenir un avantage concurrentiel grâce à l'IA générative ›

Découvrez en quoi AWS facilite la création, la mise à l'échelle et l'obtention de la valeur opérationnelle de l'IA générative ›

GLOSSAIRE



L'IA générative est un type d'IA qui peut créer de nouveaux contenus tels que du texte, du code, des images et des vidéos à partir de modèles qu'elle a appris lors d'un entraînement sur des données publiques approfondies avec des techniques d'AA.



Les modèles de fondation (FM) sont des modèles d'apprentissage profond entraînés à partir de grandes quantités de données non structurées et non étiquetées qui peuvent être utilisées pour un large éventail de tâches prêtes à l'emploi ou adaptées à des tâches particulières par la personnalisation.



Les grands modèles de langage (LLM) constituent une catégorie de FM qui peuvent traiter des quantités massives de texte non structuré et apprendre les relations qui existent entre des mots ou des portions de mots. Cela permet aux LLM de générer du texte en langage naturel, en effectuant des tâches telles que la synthèse ou l'extraction de connaissances.



La génération augmentée de récupération (RAG) est le processus qui permet d'accroître la connaissance contextuelle des LLM et de réduire les erreurs factuelles, ou hallucinations, en permettant d'augmenter les requêtes avec des données contextuelles et propres au domaine, par exemple au moyen de bases de données vectorielles.



La personnalisation est le processus par lequel un FM préalablement entraîné suit une courte période d'entraînement sur un petit jeu de données étiqueté dans le but de l'adapter pour qu'il accomplisse mieux une certaine tâche. Cet entraînement supplémentaire permet au modèle d'apprendre les nuances, la terminologie et les schémas particuliers qui se trouvent dans le petit jeu de données et de s'y adapter.



L'ingénierie de requête s'entend du processus de conception, de réglage fin et d'optimisation des requêtes d'entrée afin d'orienter un FM vers la production des résultats souhaités (soit des résultats exacts).