



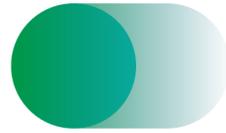
Qlik

Du Big Data au Better Data

Tendances

Vers une IA générative de confiance

2024



Il semblerait que l'engouement pour l'IA ait atteint son paroxysme, avec comme fer de lance l'IA générative, promettant de nous faire entrer dans une nouvelle ère de productivité et de prospérité.

Mais des doutes émergent, remettant en question cette promesse, pouvant aller jusqu'à l'anéantir. Jusqu'à présent, l'accès aux données pour l'IA générative s'effectuait librement, sans traçabilité de l'origine des données ni contrôle de leur qualité. Des données erronées en découlaient et venaient altérer la fiabilité des résultats faisant apparaître hallucinations, incohérences ou même contre-vérités. Laissées telles quelles, ces données « fausses », renforcées par l'IA générative, constitueront un danger exponentiel pour les entreprises et la société.

Il nous faut donc trouver un nouveau modèle encourageant l'utilisation de données de meilleure qualité et fiables : les Better Data*, c'est-à-dire des données dont la validité et la valeur ont été prouvées. Des données fiables, combinées à l'analytique et à l'automatisation, constitueront le socle qui aidera les individus et les organisations à une prise de décisions

efficace et plus avisée, tout en alimentant une IA fiable et responsable.

À terme, la fiabilité des données accélérera leur utilisation en tant que capital, les plaçant au même rang que le capital humain et le capital financier. Imaginez un produit fondamental pour l'IA pouvant être commercialisé et gagnant en valeur à mesure qu'il est utilisé.

Cette perspective est exaltante, mais comment y parvenir ? À l'heure actuelle, cette question fait débat. Nous avons passé au crible le buzz pour identifier 10 tendances en matière d'IA, de données, d'analytique et d'automatisation qui aideront les organisations à garantir que toutes leurs données sont fiables et sources de valeur au sein de l'économie de l'IA.



Dan Sommer, Senior Director, Market Intelligence Lead



* Définition de « Better Data » : données qui étendent le périmètre du Big Data en amont de l'IA, en gérant le volume, la vitesse, la variété, la validité et la valeur.

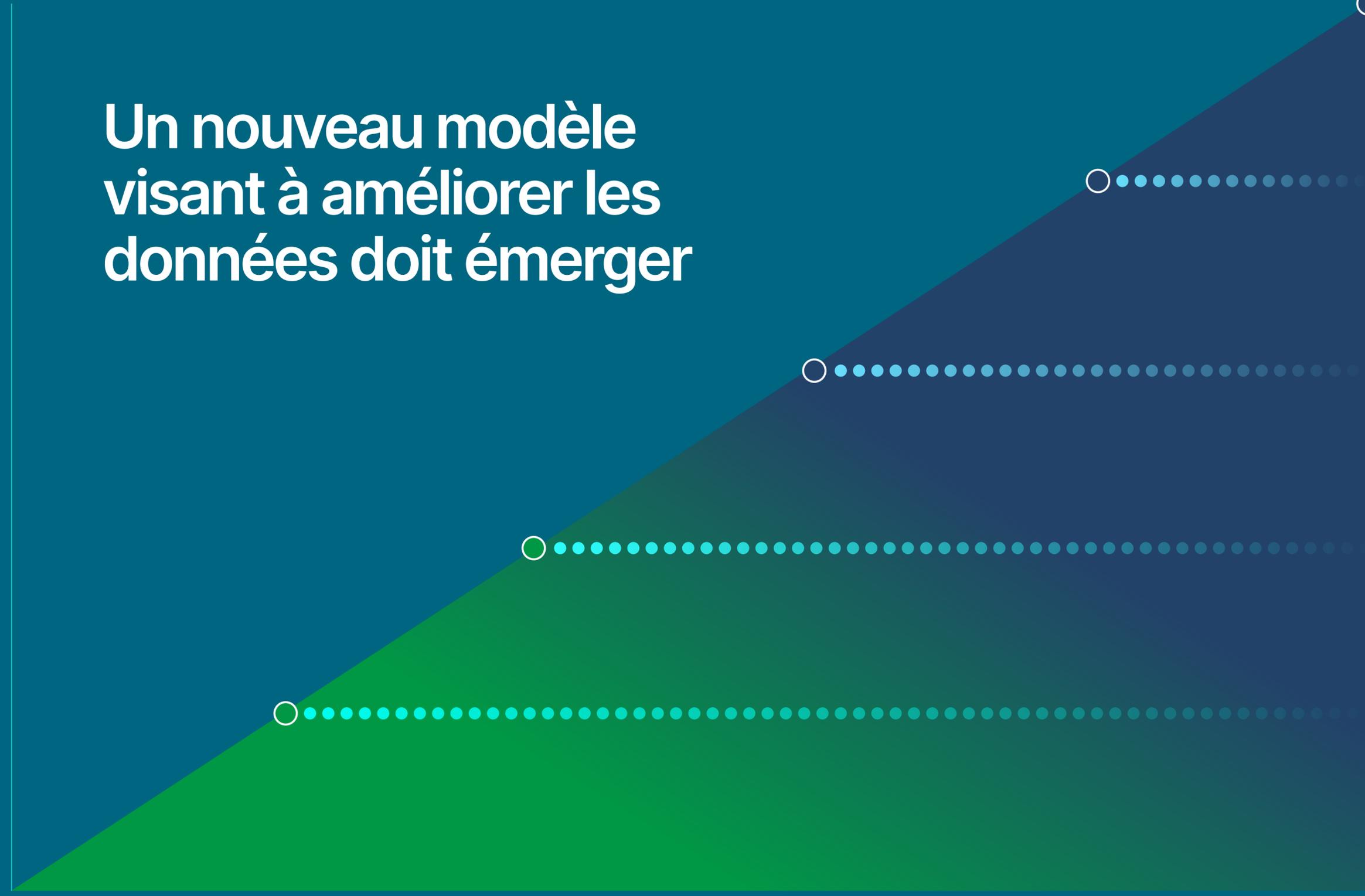
Du Big Data au Better Data : le pouvoir des 5 V

Passer du Big Data au Better Data est une progression dans la chaîne de valeur des cinq V : **Volume, Vitesse, Variété, Validité et Valeur.**

Le Big Data a souvent été associé aux trois V : **Volume, Vitesse et Variété.** En effet, le Big Data pourrait être défini comme une importante quantité de données qui s'accumulent souvent rapidement à partir de sources variées (ce dernier point étant un objectif que de nombreuses entreprises tentent encore

d'atteindre). Les « données fiables » ou « Better Data » vont plus loin. Elles englobent les deux autres V (sans doute les plus importants) : **Validité et Valeur.** Des données fiables ont été collectées à partir de diverses sources, leur crédibilité a été vérifiée et leur traçabilité est possible. Ces données peuvent ensuite être affinées, regroupées et gouvernées, c'est-à-dire traitées comme un produit connu et de qualité pouvant être partagé et commercialisé.

Un nouveau modèle visant à améliorer les données doit émerger



Valeur

- La data, produit commercialisable
- Le cercle vertueux de l'automatisation et de l'IA
- L'indispensable personnalisation de l'IA en bout de chaîne pour les entreprises

Validité

- L'importance de l'origine des données
- L'IA Literacy pour palier au nombre croissant de développeurs néophytes
- Fusion du data engineering, de l'analytics et de la data science

Variété

- De la BI à l'IA et inversement
- Bienvenue dans l'ère des données non structurées

Vélocité

- L'IA hybride, un bon compromis (à utiliser dès à présent)
- Booster l'expérience du consommateur data

Volume

Le volume étant un des 5 Vs le plus avancé, il n'y a pas de tendance majeure à noter.

La promesse de l'IA générative

Gartner® place l'IA générative au sommet de son Hype Cycle. En moins d'un an, elle a envahi le marché et est d'ores et déjà en train de transformer le monde.



McKinsey

2,6 à 4,4 billions USD¹

L'IA générative pourrait chaque année ajouter l'équivalent de 2,6 billions à 4,4 billions USD par an à l'économie mondiale. Dans 63 cas d'usage, cette technologie peut prendre en charge des défis métier spécifiques à travers 16 fonctions métier.

Cette promesse deviendra-t-elle un jour réalité ?

Sous sa forme actuelle, peut-elle réellement atteindre les sommets escomptés ? Jusqu'à présent, elle était « gratuite pour tous ». Les grands modèles de langage (LLM) ont pu amasser d'énormes quantités d'informations, sans garde-fous, afin d'entraîner leurs modèles. Cela a suscité des réactions négatives. De nombreuses personnes affirment que l'IA générative ne fera qu'empirer les choses, et non l'inverse.

Pourquoi ? Simplement parce que la qualité de l'IA dépend des données qui l'alimentent. À l'ère de l'augmentation exponentielle du Big Data, nous avons résolu le problème du **volume**, de la **vitesse** et travaillons sur la **variété**, nous n'avons pas encore réglé la question de la **validité** et de la **valeur**. À mesure que l'IA prendra une place centrale, la notion de GIGO (garbage in, garbage out) ne sera plus un adage, mais un réel avertissement accompagné de conséquences majeures. Nous devons tous y prêter une attention particulière, car la voie de la confiance est semée d'embûches.

¹ McKinsey: The economic potential of generative AI: The next productivity frontier, Jun 14, 2023

Le dilemme de la fiabilité des données

TROP de données

Actuellement, l'IA générative est entraînée à partir de toute expression humaine (bonne ou mauvaise), ce qui a pour conséquences de la parasiter et de la biaiser. De « bonnes » données peuvent être polluées par de « mauvaises », parfois intentionnellement. Cela peut conduire à des hallucinations et à des erreurs factuelles. Les experts prédisent que d'ici 2025, **jusqu'à 90 % du contenu en ligne sera généré par l'IA.**² Même si cela peut sembler être une excellente nouvelle en termes de productivité, ce chiffre est alarmant quant à la précision du contenu. Si la proportion de données synthétiques (fausses) par rapport aux données fiables devient trop élevée, cela engendrera une « dérive de la réalité », c'est-à-dire un résultat dans lequel l'information est déconnectée du monde réel.

PAS ASSEZ de données

Dans le même temps, des organisations dissimulent des données, limitant l'étendue des connaissances disponibles pour les LLM. Les géants des réseaux sociaux et les conglomérats médiatiques, dont **Reuters et CNN**, ont entamé une rébellion contre les données, limitant celles présentes sur leurs systèmes pour en restreindre l'accès. Cette décision réduit la quantité d'informations de qualité à disposition de l'IA générative, affectant négativement les résultats, et pourrait marquer le début d'une économie des « données pour l'IA ».

Environnement réglementaire

La confiance étant la clé de toute adoption technologique, les régulateurs cherchent à freiner la propagation de la désinformation et à appliquer des contrôles de confidentialité. Certains pays, comme l'Italie, ont adopté une approche radicale et ont **interdit les outils d'IA générative tels que ChatGPT**. La pression pour une meilleure gouvernance, une origine plus fiable et une traçabilité plus claire des données ont également conduit à l'édiction de lois. L'AI Act de l'UE, actuellement en cours de révision, considère également les LLM comme étant « à haut risque » et soumis à de multiples restrictions. Tous ces efforts pourraient ralentir, voire mettre un terme aux modèles existants, ce qui entraverait l'efficacité de l'IA générative.

Calcul et développement durable

Les besoins en puissance de calcul de l'IA générative sont élevés, ce qui entraîne une augmentation des coûts et un impact indésirable sur l'environnement. On estime que **les data centers dédiés à l'IA pourraient consommer plus d'électricité que l'ensemble des Pays-Bas d'ici 2027.**³ Si rien n'est fait, l'économie mondiale et la dynamique en matière de développement durable seront affectées.

² Nina Schick entretien Yahoo Finance Live, 7 janvier 2023

³ Alex De Vries, a PhD candidate at the VU Amsterdam School of Business and Economics, première publication dans Joule, 10 octobre 2023

Dépendance humaine

Le discernement des utilisateurs et l'étiquetage manuel sont toujours nécessaires pour l'IA générative. Toutefois, tandis que les innovateurs tentent d'équilibrer surveillance et efficacité, le manque de compétences adaptées au sein de la main-d'œuvre pourrait ralentir les progrès.

Échecs cuisants d'entreprises

Nous n'avons pas encore observé d'échecs largement médiatisés de l'IA générative mettant des entreprises dans une situation délicate, mais ils sont susceptibles d'apparaître à mesure que son utilisation augmente. Tous les regards étant tournés vers ceux qui sont les premiers à adopter cette technologie, un faux pas pourrait avoir de graves conséquences.

La crise de la confiance

Les entreprises demeurent extrêmement méfiantes à l'égard de l'IA générative. Les données de Qlik indiquent que **seuls 39 % des organisations⁴** ont une stratégie d'IA formalisée en place. Et cela se reflète dans les politiques d'entreprise. Les grandes entreprises technologiques hésitent par exemple à ce que leurs employés utilisent ChatGPT, car ce type d'outil est encore loin d'être adapté aux besoins de l'entreprise. Il est toutefois crucial que les données des entreprises soient adaptées à l'IA. Y parvenir nécessite de se pencher sur deux des cinq V : Validité et Valeur.

L'importance d'un nouveau modèle dès aujourd'hui

Il est dangereux de déployer l'IA générative sans disposer de données de qualité, gouvernées et fiables. Un nouveau cadre, à la fois à l'intérieur et à l'extérieur des organisations, doit être mis en place, imposant davantage de responsabilité quant à la qualité, à la traçabilité et à la transparence des données. Ce processus doit avoir lieu le plus tôt possible, mais les conséquences négatives ne doivent pas être l'unique moteur de ce changement. Il doit également utiliser des mesures incitatives, en récompensant la création de produits data de qualité, afin de démocratiser leur utilisation.

On connaît la chanson...

Au cours des dernières décennies, la musique est devenue gratuite et accessible à tous sur des plateformes telles que Napster et Pirate Bay. Cependant, ce modèle n'était pas viable. L'essor des plateformes de streaming a modifié le modèle économique, permettant une redistribution gouvernée et alignée sur la gestion des droits. Cette approche a à son tour profité aux titres les plus écoutés, ainsi qu'à leurs auteurs et créateurs.

C'est précisément le type d'évolution qui doit avoir lieu dans l'économie de la data afin d'alimenter correctement l'IA. Les créateurs de produits data de qualité doivent disposer de davantage de lieux d'échange pour les commercialiser. Ainsi, fournir des informations valides et fiables dont les origines sont traçables aura un intérêt financier sous-jacent. Un autre aspect important est que plus les données sont utilisées et valorisées, plus leur qualité est renforcée.

La gestion, l'analytique et l'automatisation des données seront autant de facteurs clés qui apporteront efficacité et **valeur** à l'IA. **Mais quelle est la roadmap de ce modèle en action ? Voici 10 tendances de l'IA qui auront un impact sur les organisations en 2024.**



1

L'IA hybride, un bon compromis

L'IA traditionnelle parviendra-t-elle un jour à maturité ou l'IA générative prendra-t-elle sa place ?

Actuellement, l'IA générative et son vaste potentiel sont sur toutes les lèvres. Toutefois, plusieurs autres initiatives en matière d'IA se développent depuis des années et certaines commencent à porter leurs fruits. Il s'agit par exemple du machine learning (ML), qui se démocratise, ou des augmentations générales tout au long du pipeline. L'une des principales idées fausses à propos de l'IA générative est qu'elle remplacera tout le reste. C'est loin d'être vrai. Le BARC a déclaré l'année dernière « Fini de jouer » avec les initiatives « traditionnelles » d'IA, ce qui signifie que l'IA a suffisamment mûri pour être mise en production et à l'échelle. Cela est particulièrement vrai pour des cas d'usage bien établis tels que l'analyse des fraudes et des taux d'attrition. Ainsi, tandis que l'IA générative fait sa place, le machine learning et d'autres formes d'IA ont déjà montré un potentiel illimité et pourraient même être utilisés pour créer le bon compromis.

Prédiction
des analystes

~30%

L'IA générative devrait atteindre environ 30 % du marché global de l'IA d'ici 2025.⁵

⁵ Boston Consulting Group

2

L'IA générative, génératrice d'informations : booster l'expérience du consommateur data

Comment l'IA autonomise-t-elle les individus disposant de moins de compétences techniques ?

Tout le monde ne souhaite pas créer des applications. En réalité, **les trois-quarts**⁶ des gens ne se rendent même pas compte qu'ils utilisent des outils d'analyse, de près ou de loin. Tout ce que souhaitent ces professionnels de l'information « ordinaires » est une réponse, idéalement immédiate, mais ils n'ont ni le temps, ni l'envie, ni les compétences nécessaires pour effectuer une analyse. Autre constat : les consommateurs ont tendance à faire confiance aux individus plutôt qu'à la data. La collaboration et le partage de données sont donc essentiels. Cette base d'utilisateurs adore les visualisations et les informations générées automatiquement, enrichies d'explications en langage naturel. Le fait que cela soit possible au sein des systèmes dans lesquels ils opèrent est un atout supplémentaire.

Prédiction
des analystes

66%

D'ici 2025, 66 % des entreprises du G2000 adopteront une BI et une analytique « sans tête » basées sur l'IA avec des fonctionnalités de chat, de questions-réponses et de notifications proactives, multipliant ainsi par 4 le nombre d'utilisateurs ayant accès aux informations contextuelles.⁷

⁶ Enquête BARC (Business Application Research Center) et Eckerson Group « Strategies for Driving Adoption and Usage with BI and Analytics », mars 2022

⁷ IDC FutureScape: Worldwide Data and Analytics 2024 Predictions, IDC #US51295223, octobre 2023

3

Voici venue l'ère des données non structurées

L'IA générative est-elle la clé pour faciliter l'analyse des données non structurées ?

La majorité (**Forrester parle de 80 %**) des données dans le monde ne sont pas structurées. En d'autres termes, elles ne sont pas soigneusement triées en lignes et colonnes.

Les e-mails et les documents sur votre intranet en sont un exemple. Nombreux sont ceux qui ont déjà tenté, sans succès, d'analyser des données non structurées. Cela est désormais possible grâce aux nouvelles métadonnées et techniques sémantiques. En utilisant des graphiques de connaissances et des bases de données vectorielles, complétés par la technique RAG (Retrait, Augmentation, Génération), l'association de données structurées et non structurées de manière fiable présente d'innombrables possibilités. Avec une couche de gestion des réponses, vous pouvez réutiliser des questions et réponses vérifiées et fiables, ce qui vous permet d'analyser l'intégralité de votre patrimoine de données et d'utiliser des LLM privés créés en interne grâce à l'analyse des données.

Prédiction des analystes

2x

Le volume des données non structurées gérées par les entreprises doublera au cours de 2024.⁸

⁸ Forrester, Predictions 2024: Data And Analytics

4

De la BI à l'IA et inversement : l'évolution de l'analyse métier

Comment l'IA générative prend-elle en charge l'analyse métier ?

La révolution de l'IA générative avance à une vitesse fulgurante, ouvrant la voie à de nouvelles façons d'interagir avec les données, notamment l'analytique multimodale. Vous pouvez désormais simplement faire glisser un fichier sur une interface de chat et commencer la discussion. Cette interface peut alors générer des requêtes et du code, aider à créer du contenu et accélérer les processus automatisés. Pour un nombre croissant de personnes, le parcours analytique commence avec ces outils d'IA générative. Elles les utilisent pour obtenir des visualisations de données simples et des projections métier. Ici, la BI se sert de l'IA.

Dans un deuxième temps, ces individus voudront peut-être utiliser des outils d'entreprise pour effectuer une analyse plus approfondie, apportant ainsi les avantages de l'IA générative aux outils auxquels ils font confiance. L'IA se sert alors de la BI. En d'autres termes, nous passons d'un mode à l'autre (via l'intégrabilité, la connectivité et les API) pour tirer le maximum d'avantages de chaque plateforme.

Prédiction des analystes

70%

Selon Gartner®, d'ici 2026, l'IA générative modifiera considérablement 70 % des initiatives de conception et de développement de nouvelles applications web et mobiles.⁹

⁹ Gartner, Top Strategic Technology Trends for 2024, 16 octobre 2023. GARTNER est une marque déposée et une marque de service de Gartner, Inc. et/ou de ses affiliés et est utilisée ici avec autorisation. Tous droits réservés.

5

L'importance de l'origine des données : comprendre leur ADN

Comment se fier à des données dont on ignore l'origine ?

Comme nous l'avons vu précédemment, la qualité et la traçabilité des données ont toujours été importantes. Aujourd'hui, avec l'omniprésence de l'IA dans notre quotidien, elles sont devenues incontournables. Cette exigence est essentielle pour les données nécessaires au bon fonctionnement de votre entreprise et à l'entraînement des modèles d'IA. La nécessité d'identifier et de comprendre l'origine des données est particulièrement importante dans le cas des LLM publics, dont l'origine n'est actuellement pas traçable. Sans ces informations, il est difficile pour les meilleurs modèles d'IA générative de différencier la réalité de la fiction. Cela peut entraîner des hallucinations, des faits erronés et des deepfakes.

Pour les entreprises, faire confiance à de tels résultats peut avoir de graves conséquences. C'est pourquoi les organisations doivent en faire une priorité dès maintenant.

Un mécanisme permettant d'étiqueter et d'identifier clairement les données est nécessaire, mettant à profit des techniques de provenance et de cryptographie, ainsi que des méthodes qui restent à inventer afin de créer ce qu'on pourrait appeler un « test ADN des données ». Plusieurs initiatives sont déjà en cours, comme **The Coalition for Content Provenance and Authenticity** dont les membres ne sont autres qu'Intel, BBC et Sony ;

le filigrane de Google (SynthID) qui identifie les images générées par IA et **la carte de modèle de Hugging Face** qui crée de simples fichiers Markdown avec des métadonnées supplémentaires.

La confiance dans l'origine et la traçabilité des données entraîne un cycle qui se perpétue automatiquement, dans lequel les individus assument la responsabilité des données. Il s'agit également de l'une des pièces les plus importantes du puzzle permettant de transformer les données exclusives de l'entreprise en produits commercialisables et filigranés.

Prédiction
des analystes

90%

du contenu en ligne sera
généralisé par l'IA d'ici 2025 ¹⁰

¹⁰ Nina Schick entretien Yahoo Finance Live, 7 janvier 2023

6

L'IA Literacy pour palier au nombre croissant de développeurs néophytes

Comment mettre cet immense pouvoir entre les mains des développeurs citoyens, en toute sécurité ?

En peu de temps, le langage naturel a éclipsé le low code, comme nouveau langage de programmation dominant. Lorsque le codage est simplifié, des tâches plus avancées comme la création d'applications en sont facilitées. Ce processus entraînera une explosion du nombre d'applications créées par des « développeurs amateurs ». Résultats ? L'arrivée d'un tourbillon d'innovations, mais également un risque de créer le chaos en matière de gouvernance et de voir apparaître une profusion d'applications. Alors que ce processus confère des pouvoirs très puissants au plus grand nombre, les organisations doivent prendre des mesures pour sensibiliser leur personnel aux avantages et aux pièges de l'IA générative. Si les cinq dernières années ont été consacrées à l'enseignement de la Data Literacy aux équipes, nous devons maintenant nous tourner vers l'AI Literacy. En parallèle, la gestion du cycle de vie des applications et la promotion des données et applications pertinentes revêtiront une nouvelle importance.

Prédiction des analystes

60%

D'ici la fin 2025, pour atténuer les nouveaux risques provenant de l'utilisation généralisée de l'IA générative par leur personnel, 60 % des grandes entreprises imposeront une Data Literacy formelle et une formation à l'IA responsable.¹¹

¹¹ IDC FutureScape: Worldwide Future of Enterprise Intelligence, Prédiction 2024, IDC #US51293423, octobre 2023

7

Fusion du data engineering, de l'analytics et de la data science

Le fait de ne plus avoir un besoin impératif de connaissances et d'outils avancés pour traiter les données entraînera-t-il leur démocratisation ?

Selon IDC¹², les entreprises préfèrent travailler sur l'ensemble du pipeline de données avec des fonctionnalités de pointe en faisant appel à moins de fournisseurs, voire un seul. Les nouvelles plateformes, combinées à l'évolution des data fabrics, « consommeront » l'ingénierie des données pour une nouvelle génération d'utilisateurs, tout particulièrement si elles s'accompagnent d'une IA, d'une automatisation et d'une data science puissantes. Cela permettra par exemple aux analystes métier de revenir en arrière plus tôt dans le pipeline pour effectuer des tâches de gestion et de préparation des données. Ces mêmes analystes pourront également appliquer des modèles statistiques avancés aux données et aux outils avec lesquels ils travaillent quotidiennement, sans avoir besoin de les exporter vers un workbench avancé.

¹² Enquête IDC Data Management, 2023

¹³ Gartner, Predicts 2023: Analytics, BI and Data Science Composability and Consolidation. GARTNER est une marque déposée et une marque de service de Gartner, Inc. et/ou de ses affiliés et est utilisée ici avec autorisation. Tous droits réservés.

Simplifier les tâches difficiles, ainsi que fusionner les rôles et les capacités de l'ingénierie des données, de la data science et de l'analyse, permettra aux organisations de résoudre des problématiques plus complexes. Nous passerons de la question « Quel sera le montant des bénéfices réalisés ce trimestre ? » aux questions « Quels clients devrions-nous cibler à l'avenir ? » et « Parmi nos meilleurs employés, quels sont ceux les plus susceptibles de partir et par quels les facteurs spécifiques cette décision est-elle motivée ? ». Rapprocher des fonctions auparavant cloisonnées aidera les entreprises à transformer leur Big Data en Better Data.

Prédiction des analystes

50%

Gartner® prédit que d'ici 2026, 50 % des organisations devront évaluer les plateformes ABI et DSML en tant que plateforme composable tout-en-un en raison de la convergence du marché.¹³

8

Le cercle vertueux de l'automatisation et de l'IA

Pourquoi le champ d'action de l'IA évolue-t-il pour passer de l'analyse à l'exécution ?

Jusqu'à présent, les LLM et l'IA générative ont été principalement utilisés pour appuyer les raisonnements et effectuer des analyses, plutôt que pour l'iPaaS et le passage à l'action. Cependant, plusieurs initiatives passionnantes sont en cours pour soutenir ces dernières, notamment une approche des LLM impliquant une synergie entre le raisonnement et l'action. Bien entendu, cela nécessite des données transformées en temps quasi réel au bon emplacement. Les organisations commencent à mettre en place de nouvelles façons d'utiliser l'IA générative avec l'automatisation des applications, comme l'utilisation de l'analyse des sentiments pour automatiser et générer différentes réponses en fonction de l'humeur. L'association de l'IA générative et de l'automatisation limitera les tâches manuelles visant à créer et relier des workflows, et laissera aux utilisateurs du temps pour l'arbitrage des décisions.

Cependant, parallèlement à ces nouvelles opportunités, il est crucial de réaliser que relier l'automatisation à l'IA augmentera davantage le pouvoir et les capacités de l'IA. Nous devons nous y préparer et mettre en place des garde-fous.

Prédiction des analystes

20%

Selon Gartner®, d'ici 2027, la détection des valeurs aberrantes et d'autres fonctionnalités d'analytique augmentée évolueront vers des plateformes d'analyse autonomes qui gèreront et exécuteront entièrement 20 % des processus métier.¹⁴

¹⁴ Gartner, Predicts 2023: Analytics, BI and Data Science Composability and Consolidation. GARTNER est une marque déposée et une marque de service de Gartner, Inc. et/ou de ses affiliés et est utilisée ici avec autorisation. Tous droits réservés.

9

L'indispensable personnalisation de l'IA en bout de chaîne pour les entreprises

Comment développer des outils d'IA générative pour les besoins spécifiques des entreprises ?

Pour le moment, les premières applications d'IA générative sont des projets extrêmement évolutifs, bien que toujours génériques, capables d'utiliser les LLM. Celles-ci se font généralement dans un contexte business-to-consumer (B2C). Au fil du temps, l'IA sera de plus en plus adaptée à l'industrie et à des cas d'usage business-to-business (B2B) plus spécifiques. Cela prendra la forme d'applications et de LLM privés où le socle peut être commun, mais où des couches de personnalisation permettront de mieux servir la « longue traîne ». Un exemple de ce processus est un **cluster d'IA que Mark Zuckerberg développe pour la recherche médicale.**

¹⁵ IDC FutureScape: Worldwide Future of Enterprise Intelligence, Prédiction 2024, IDC #US51293423, octobre 2023

En extrapolant cette tendance, des applications sophistiquées seront créées pour répondre à un secteur ou à un problème spécifique, économisant des efforts, et du temps consacré à la consultation ou au recueil de conseils. Les données propriétaires de votre organisation constitueront ici une matière première précieuse et des « solution fabrics » verront le jour là où les données et applications spécifiques à un domaine pourront être partagées et commercialisées. Toutefois, la question de savoir quelle IA servira de socle à cette construction demeure sans réponse. Bien qu'il existe actuellement 2 à 4 LLM dominants en cours de formation, **certain pensent que ce système s'appuiera de plus en plus sur l'open source.**

Prédiction des analystes

80%

D'ici 2026, plus de 80 % des cas d'usage de l'IA générative dans les entreprises exploiteront des modèles d'IA personnalisés et spécialisés plutôt que des modèles de base génériques proposés via des API publiques.¹⁵

La data, produit commercialisable

Comment l'IA est-elle devenue un élément critique de la monétisation des données de votre organisation ?

Les approches architecturales visant à harmoniser les données distribuées et variées, telles que les data fabrics et data meshes, se sont concrétisées au cours de l'année dernière grâce aux avancées technologiques et dans le domaine de l'IA. L'un des éléments clés de ces approches, qui trouve un écho auprès des clients, est la « Data-as-a-Product ». Il s'agit d'appliquer les principes de gestion de produits aux données, en se posant des questions sur les problématiques que nous résolvons, et en se demandant qui utilisera les données et pour quoi. Ce processus souligne l'importance de la qualité, de la gouvernance et de la facilité d'utilisation des données pour les utilisateurs finaux. La data, considérée comme un produit, évolue pour devenir le socle de la consommabilité de toutes les formes d'analytique et d'IA.

Le concept de traitement des données en tant qu'actif ou produit de valeur signifie qu'elles peuvent figurer dans un catalogue, être utilisées à diverses fins en interne et même devenir un bien commercialisable. L'objectif est de monétiser la data, considérée alors comme un produit, hors de votre organisation. Nous commençons à observer davantage de plateformes où les données validées peuvent être affinées, achetées, vendues et échangées, avec une rémunération à la clé pour ceux qui les détiennent. Le lancement récent des « GPT » par OpenAI constitue une étape importante et un tournant décisif, car il s'agit d'une approche « app-store » pour les applications d'IA contextualisées, avec un modèle de partage des revenus. Ultérieurement, elles seront enrichies avec

des données supplémentaires. Cela devrait encourager les organisations à utiliser leurs propres données pour perfectionner les modèles de ChatGPT, qui pourront ensuite être monétisés. À l'avenir, des échanges similaires serviront de sources vérifiées à partir desquelles les LLM pourront explorer des données validées et distribuer une compensation pour leur accès, comme l'a fait l'industrie de la musique avec les services de streaming. Plus le produit data est utilisé, plus il prend de la valeur.

Prédiction des analystes

60%

D'ici 2026, 60 % des principales sociétés de BI auront identifié des produits data, et 15 % auront attribué une valeur métier à ces produits grâce à une méthodologie de valorisation des données.¹⁶

¹⁶ IDC FutureScape: Worldwide Future of Enterprise Intelligence, Prédiction 2024, IDC #US51293423, octobre 2023

4e et 5e Vs

Validité et Valeur, deux éléments cruciaux pour l'avenir
de l'IA générative

Conclusion

La qualité et la traçabilité des données ont toujours été importantes. Aujourd'hui, avec l'omniprésence de l'IA dans notre quotidien, elles sont devenues incontournables. Et c'est pour cette raison que les 5 V doivent être présents. Nous avons résolu le problème du volume et de la vitesse et travaillons encore sur la variété. Si nous voulons vraiment obtenir des données qui tiendront la promesse de l'IA générative, nous devons passer du Big Data au Better Data, c'est-à-dire à des données fiables. Cela nécessite de régler la question de la validité et de la valeur. Les données et métadonnées de votre entreprise constituent un atout unique. L'ajout des deux autres V, Validité et Valeur, est la garantie de faire de la data un asset exploitable pour l'IA.

L'avenir promis par l'IA générative repose sur un seul et unique élément : la qualité des données utilisées par cette technologie. Si l'origine et la qualité des données sont vérifiées de manière cohérente et approfondie, elles peuvent être transformées en produit. En outre, plus vos données sont utilisées par l'IA, plus elles prennent de la valeur, tant en interne qu'en externe. Nous assisterons à une évolution où le Better Data deviendra la matière première alimentant les LLM de confiance et deviendra un bien commercialisable. Le capital data prendra de l'importance et sera à la base de toute innovation utilisant l'IA générative.

Bien entendu, reste à savoir comment obtenir des données de meilleure qualité, générant une plus grande valeur. Toutefois, ce qui est considéré comme ayant de la valeur à l'ère de l'IA générative pourrait vous surprendre. Revenons à l'analogie classique de la ruée vers l'or : ceux qui en ont massivement profité sont ceux qui ont vendu les pelles, les jeans ou les bottes, organisé les transports et construit les chemins de fer. De même, à l'ère de l'IA, être une plateforme fiable et permettre d'extraire des données dites Better Data, optimisées par l'analytique et l'automatisation, sera une grande source de valeur.

Vous devez agir dès maintenant. Nous vivons le calme avant la tempête, avant que l'IA ne soit intégrée dans tous les aspects du travail de la connaissance. Gare à la complaisance ou vous serez laissé pour compte. L'IA générative va changer le monde dans la même mesure que l'a fait Internet. Cela ne sera pas simple. Mais, en prenant les bonnes mesures et en canalisant correctement les résultats, surmonter les obstacles ouvrira la voie à une ère d'innovation et de prospérité sans précédent.

Si vous vous sentez dépassé, n'oubliez pas que vous n'êtes pas seul. Collaborez avec des partenaires compétents qui peuvent transformer votre Big Data en Better Data, et profitez vous aussi de la valeur générée par l'IA générative.