

Tendances Data & IA | 2024



TENDANCES

Data & IA

[sf≡ir]
WENVISION

#TechTrends

Au menu de 2024 : GenAI, Qualité, Fiabilité

3 principes directeurs pour investir dans un SI data compatible avec le futur

Le monde de la data est en constante évolution, et l'arrivée fracassante de **l'intelligence artificielle générative (GenAI)** a accéléré cette transformation. Si la GenAI offre des possibilités inédites d'interaction avec les données, elle met également en lumière l'importance d'une plateforme data solide et bien structurée. **Il ne peut y avoir d'application de GenAI ou d'IA sans une plateforme de données fiable et bien gouvernée.** Dans ce contexte, trois piliers se distinguent comme les **Tech Trends Data** incontournables de 2024 : la **gouvernance** des données, la **qualité** des données et l'intégration de la **GenAI**.

- **La gouvernance des données** : englobe l'ensemble des processus et outils permettant de gérer et de contrôler les données tout au long de leur cycle de vie. Elle assure la **traçabilité, la sécurité et l'accessibilité des données**, éléments essentiels pour une utilisation optimale.
- **La qualité des données** : données erronées, incomplètes ou incohérentes peuvent fausser les analyses et les prises de décision. Avec la complexification des plateformes data, les outils de **data quality et de lineage** deviennent indispensables pour **fiabiliser** les processus et les prises de décisions.
- **L'IA Générative** : Comment passer à côté de l'une des révolutions technologiques les plus importantes de ces dernières années ? **La GenAI va profondément transformer la façon dont nous interagissons avec les systèmes d'information.** La généralisation de cette technologie est impressionnante, mais elle doit toujours être considérée comme un outil et utilisée à bon escient. Il est important de se rappeler que la GenAI est un outil, et non une solution miracle.

Bienvenue dans ces TechTrends Data , bonne lecture !



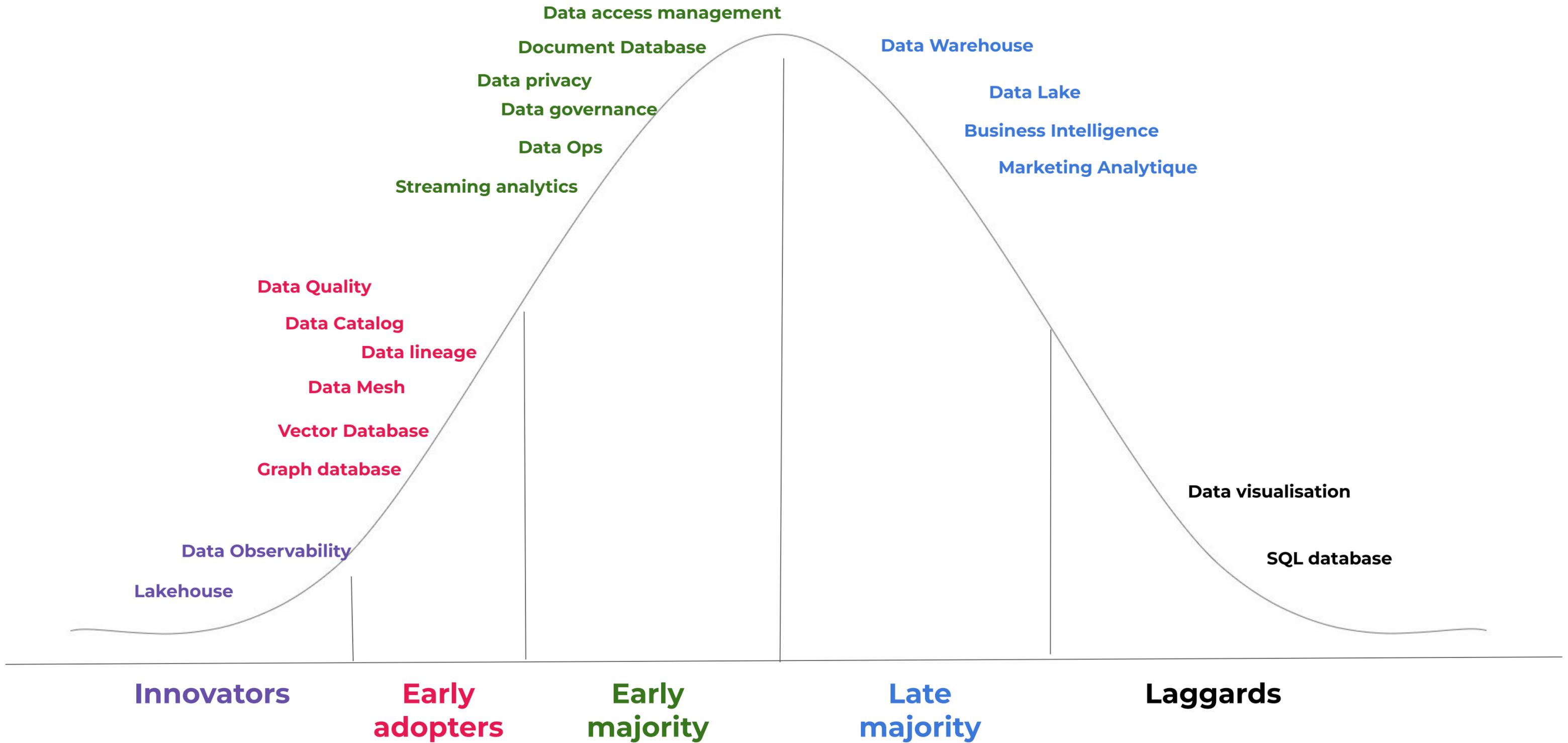
Comment interpréter notre classement

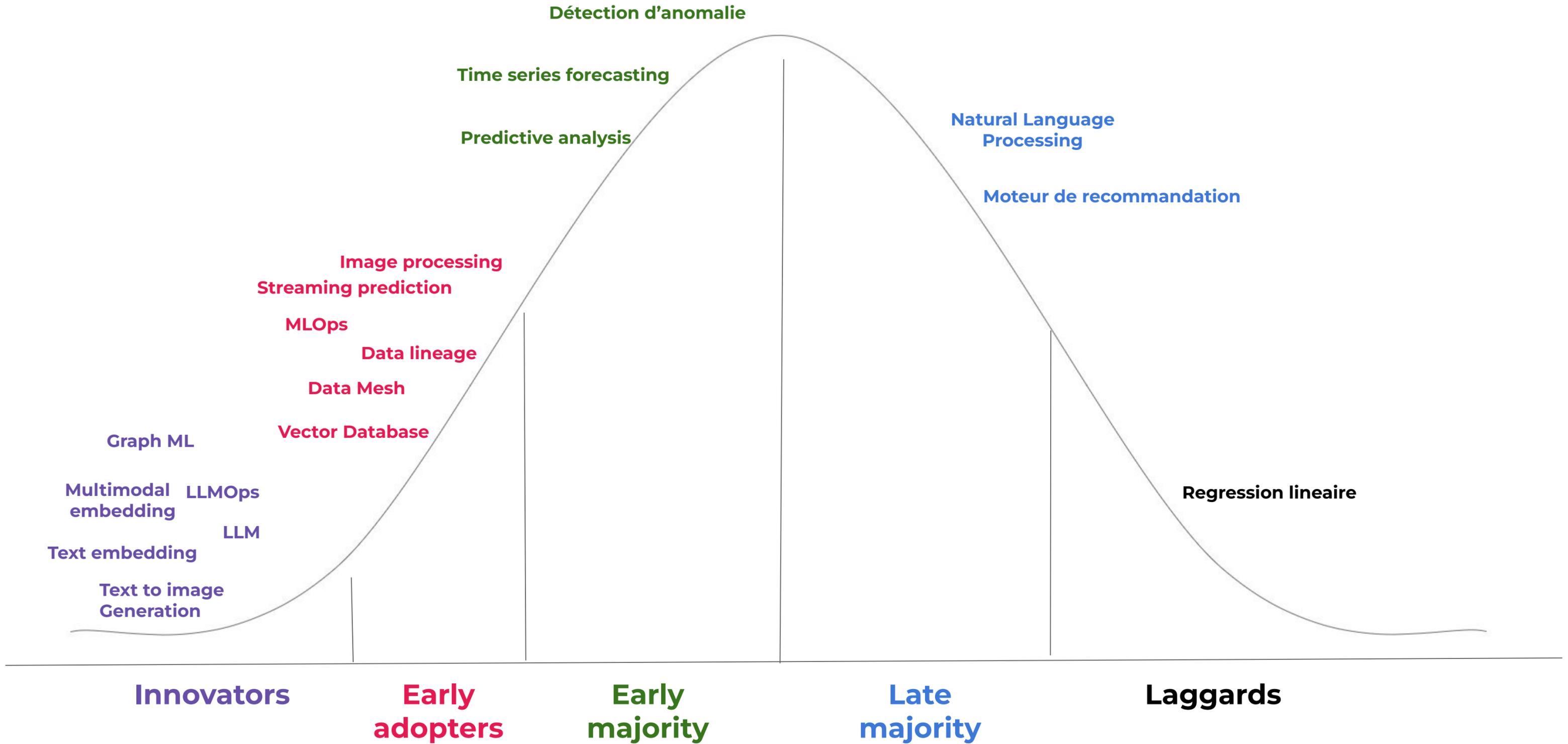


Afin d'élaborer les TechWaves 2024, nos experts Data ont pris le soin de classer les concepts technologiques, retenus dans l'étude, suivant la théorie de la diffusion des innovations d'Everett Rogers qui permet d'examiner comment les idées sont diffusées auprès de groupes d'individus.

5 catégories d'adeptes d'innovations

- **Innovators** : Ce sont les premiers à adopter les nouvelles idées. Ils sont souvent à la recherche de nouvelles idées et sont prêts à prendre des risques. Ils ont généralement des ressources financières importantes et une bonne compréhension des technologies complexes. Ils sont prêts à accepter les échecs occasionnels.
- **Early adopters** : Les adopteurs précoces adoptent les nouvelles idées peu de temps après les innovateurs. Ils sont souvent des leaders d'opinion dans leur communauté. Ils ont une bonne réputation et sont respectés par leurs pairs. Ils prennent le temps de réfléchir avant d'adopter une nouvelle idée.
- **Early majority** : La majorité précoce adopte les nouvelles idées juste avant la majorité de la population. Ils interagissent fréquemment avec leurs pairs, mais n'occupent pas souvent des postes de leadership. Ils jouent un rôle important dans le processus de diffusion des innovations. Ils prennent plus de temps que les innovateurs et les adopteurs précoces pour adopter une nouvelle idée.
- **Late majority** : La majorité tardive adopte les nouvelles idées juste après la majorité de la population. Ils peuvent être sceptiques et réticents à adopter les nouvelles idées. Ils adoptent souvent les nouvelles idées par nécessité économique ou pression sociale. Ils ont besoin de voir que l'innovation est largement adoptée avant de l'adopter eux-mêmes.
- **Laggards** : Ce sont les derniers à adopter les nouvelles idées. Ils sont souvent traditionalistes et résistants au changement. Ils ont peu d'influence sur les autres. Ils adoptent souvent les nouvelles idées lorsqu'elles sont déjà obsolètes.





01. DATA TECH

[#Stockage](#)
[#DataValorisation](#)

02. DATA ORGANISATION

[#DataConcepts](#)
[#Gouvernance](#)

03. IA

[#PredictiveAnalysis](#)

04. GEN IA

[#Culture](#)
[#Plateforme](#)
[#Gouvernance](#)

SOMMAIRE

DATA TECH



DATA TECH

LES CONTRIBUTEURS
DE CETTE SECTION

ILS ET ELLES ONT ÉCRIT LES TECH TRENDS 2024



François DENIS
Head of Cloud Advisory,
WENVISION



Ravidhu DISSANAYAKE
Consultant Data / IA,
WENVISION



Martin ELIARD
Consultant Data / IA,
WENVISION



Marie FONTAINE
Head of AI
WENVISION



Didier GIRARD
Co-CEO,
SFEIR & WENVISION



Mylène MAIGNANT
Consultante IA,
WENVISION



Seifeddin MANSRI
CTO Cloud Groupe,
SFEIR



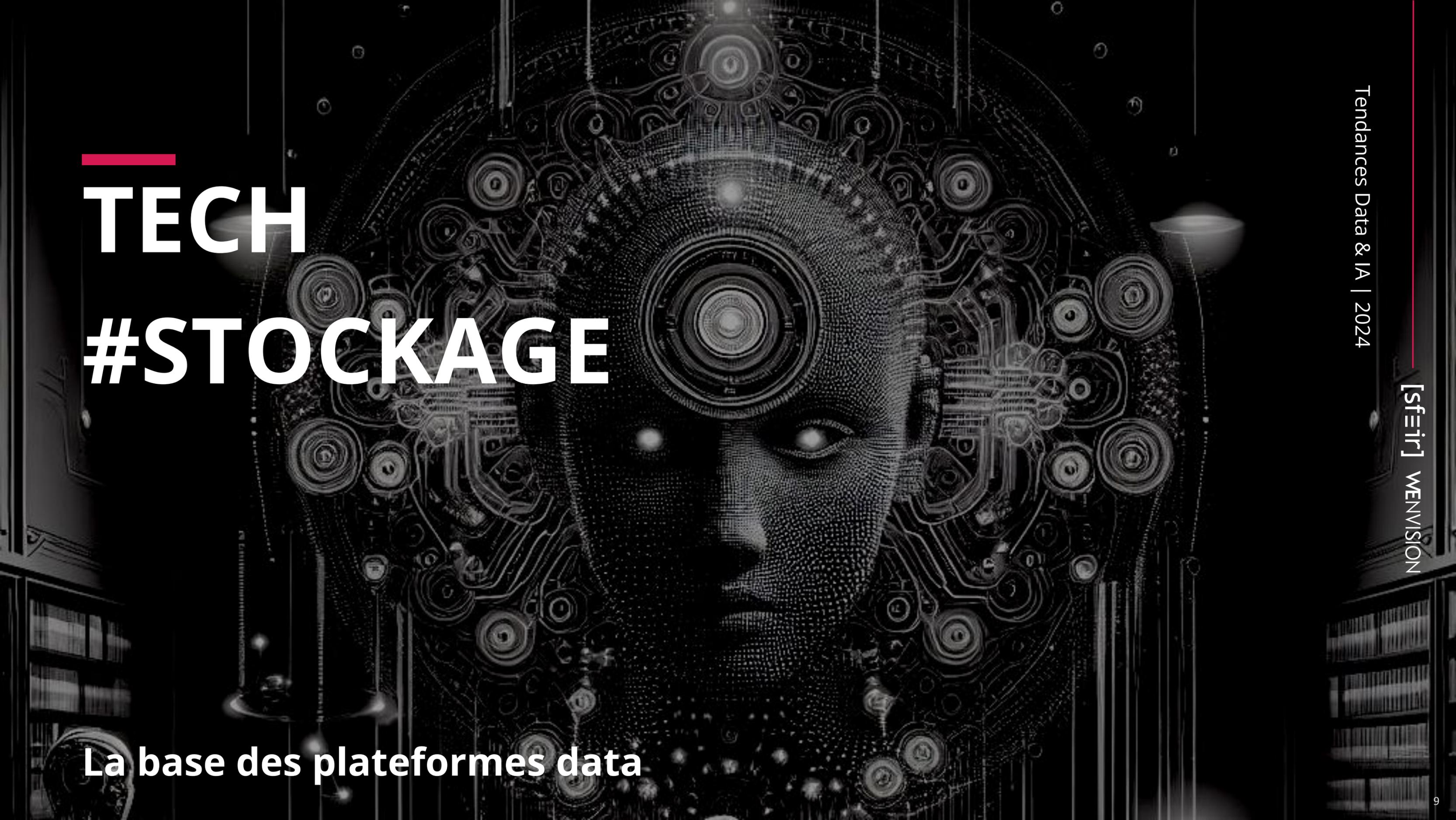
Baptiste PUGNAIRE
Consultant Data / IA,
WENVISION



Olivier RAFAL
Consulting Director Strategy,
WENVISION



Florent LEGRAS
CTO Data Groupe,
SFEIR



TECH
#STOCKAGE

La base des plateformes data

#Data #DataConcepts #Lakehouse #DataPlatform

Du Data Lakehouse à la Data Intelligent Platform

Le concept de Data Lakehouse, poussé par des acteurs comme Databricks, Starburst ou Cloudera, vise à apporter “le meilleur des deux mondes” du datalake et du datawarehouse.

Pendant longtemps, les datawarehouses étaient les seuls outils à notre disposition pour stocker et préparer les données structurées aux fins d'analyse de type BI. Les datalakes ont permis d'accueillir beaucoup plus de données, notamment des données non structurées (images, texte, vidéos...). Ils sont aujourd'hui régulièrement utilisés de manière conjointe, au sein d'une Data Platform dans le cloud, le datalake ingérant la donnée brute, dont une partie sera transformée et injectée dans le datawarehouse.

L'idée des Data Lakehouses est de s'appuyer sur un unique produit apportant l'ensemble des fonctionnalités. Toutefois, les opérations de transformation, par exemple, devront toujours exister. Data Lakehouse unique ou Data Platform utilisant conjointement deux produits ou plus, le principe est très proche et devient la pierre angulaire des infrastructures data modernes.

De la GenAI pour intégrer de l'intelligence au sein des lakehouses

Bien sûr, les Lakehouses ayant l'ambition d'agrèger des fonctionnalités et de faciliter la vie des utilisateurs, y compris des utilisateurs métiers plus loin du code et de l'informatique, ils regardent aujourd'hui du côté de l'IA générative. Les fournisseurs souhaitent notamment enrichir les capacités suivantes à l'aide de la GenAI :

- **faire du no-code et exprimer en langage naturel les requêtes et fonctions à mettre en place.** L'idée ici est de réutiliser les principes de GenAI adressés aux développeurs via une interface simple d'utilisation. Il ne sera donc plus nécessaire de connaître le langage de la plateforme ou de maîtriser l'ensemble des aspects de data engineering. Le module d'IA générative fera la traduction.
- **des modules de préparation des données intelligents.** Les data scientists n'auraient ainsi plus à passer autant de temps à nettoyer les données, adapter leur format, réaliser des contrôles de qualité. L'IA générative, avec sa capacité à traiter du contenu de la donnée, serait en capacité *a minima* de faire des propositions intelligentes, voire de réaliser les préparations et traitements en amont permettant aux data scientists de se concentrer sur leur algorithme.

Le lakehouse augmenté à l'intelligence artificielle devient donc une Data Platform intelligente et permet à un plus grand nombre d'utilisateurs de s'en saisir, sans limite de compétences techniques.



#Data #DataConcepts #BaseVectorielle

Les bases vectorielles, pour une IA générative d'entreprise



Alors que l'intelligence artificielle générative gagne du terrain dans le monde de l'entreprise, un élément technologique émerge et se distingue par son potentiel disruptif : les bases vectorielles. Véritables piliers d'une intégration sur-mesure des LLM, elles promettent une révolution aussi bien en termes de coût que de performance.

Ces bases ne sont pas nouvelles. Elles sont utilisées pour stocker des vecteurs, c'est-à-dire une représentation mathématique des contenus dans plusieurs dimensions. Elles sont de ce fait très utiles pour du clustering de données, de la recherche par similarité, de la recommandation... Cela en fait des alliées naturelles de l'IA générative en entreprise : grâce à cette technologie, il devient possible de redéfinir la gouvernance et l'exploitation des données en entreprise.

De fait, en entreprise, le véritable intérêt de l'IA générative réside non dans ses propres connaissances, mais dans sa capacité à restituer des informations ancrées dans le vocabulaire et les connaissances de l'entreprise ; on parle alors de "grounding" (ancrage). La principale technique pour cet ancrage est le RAG (retrieval augmented generation), où les documents et données de l'entreprise sont indexés sous forme de vecteurs, au travers du processus d'"embedding". Cela peut se faire à la volée, mais pour un usage du RAG sur des données persistantes, il est commode de stocker ces vecteurs dans une base ad hoc.

Des réponses pertinentes et à jour - si les sources de données sont de qualité

Les entreprises débordent de spécificités qui nécessitent de tels LLM personnalisés. Les bases vectorielles sont donc des alliées privilégiées pour ajouter du contexte dans les interactions afin que les LLM reflètent fidèlement la réalité interne de l'entreprise, sans engendrer les coûts typiquement associés à l'entraînement personnalisé de modèles.

Les "vector databases" promettent donc des analyses finement ajustées aux réalités propres de chaque entreprise. A condition que la qualité des sources soit au rendez-vous, bien sûr ; si l'IA générative s'appuie sur un portail Web qui n'a pas été mis à jour depuis plusieurs mois, par exemple, l'effet risque d'être décevant.

La question de la sûreté et de la conformité des sources de données est également primordiale. Les bases vectorielles s'avèrent de formidables garde-fous pour garantir et renforcer les politiques de gouvernance de données en contrôlant les accès et en sécurisant les informations traitées de l'entreprise. Elles permettent de plus de garantir l'efficacité des réponses fournies par les LLM (critère lié à la qualité et à la fraîcheur des données). Avec les bases vectorielles, les informations délivrées par l'IA générative peuvent être constamment mises à jour, garantissant des réponses actualisées et fiables - si tant est que les sources soient mises à jour.

#Data #DataConcepts #GraphDatabase

Les bases graphes, un atout à ne pas négliger



Face à la multiplication des données et de leurs usages, sélectionner la solution de stockage adéquate devient crucial. Si les bases de données vectorielles ont le vent en poupe grâce à l'essor des algorithmes d'intelligence artificielle générative (GenAI), les bases graphes, parfois appelées bases de connaissances, méritent une attention particulière.

Moins polyvalentes que leurs homologues vectorielles, elles ne traitent pas de données multimodales telles que des images ou de la vidéo. Les bases graphes excellent dans l'organisation et le stockage de données interconnectées. Leur particularité ? Elles ne se contentent pas de stocker des informations brutes, **mais cartographient également les relations entre les données.**

Le choix d'une base de données doit se faire en fonction de l'utilisation prévue et de la nature des données. Les bases graphes sont idéales pour déceler des schémas complexes et des liens cachés au sein des données. Déjà largement employées pour la recommandation de produits ou la détection de fraude, elles constituent également d'excellentes bases de connaissances reliant divers objets entre eux. C'est dans ce dernier domaine que l'on peut envisager un couplage des bases graphes avec les modèles de langage (LLMs) **pour fournir des informations contextuelles et pertinentes.**

En résumé, les bases graphes restent une option à considérer lors du choix d'une solution de stockage. Leur versatilité et leurs performances en font un choix judicieux lorsque les données s'y prêtent, notamment pour explorer les relations et les interconnexions.

#Data #DataConcepts #DocumentStore

Datalake, Data Warehouse, Lakehouse



Naviguer dans l'univers des solutions de stockage de données peut s'avérer complexe. Une bonne compréhension des différents systèmes et de leurs applications est essentielle pour choisir les éléments qui composeront une plateforme data efficace.

Le Data Lake : un point d'entrée flexible

Le data lake est un espace de stockage pouvant accueillir des données de différents formats et provenant de sources multiples. Il constitue souvent le point d'entrée d'une plateforme data. Sa flexibilité et sa durabilité sont ses principales caractéristiques. Généralement, les données brutes arrivent dans le data lake pour être ensuite analysées, subissant peu de modifications.

Le Data Warehouse : structure et analyse simplifiée

Pour une meilleure structuration et une analyse des données simplifiée, un data warehouse peut être intégré à la plateforme. Contrairement au data lake, il manipule des données structurées et organisées. Cette contrainte se transforme en atout lors de l'analyse (en SQL). Des données bien structurées et modélisées facilitent leur utilisation, permettant à toute personne familière avec le langage SQL d'effectuer des analyses et d'explorer les données.

Le Lakehouse : un hybride prometteur

Les Lakehouse, apparues il y a quelques années, sont des solutions hybrides. Elles permettent de stocker les données dans un stockage objet et de les interroger via des interfaces de type SQL ou Python.

Le principal avantage mis en avant par ses partisans est la possibilité d'analyser les données sans les copier dans une autre structure. Analyse et stockage se font au même endroit. La distinction entre Lakehouse et Data Warehouse ten

#Data #DataConcepts #Langages

Le SQL éclipsé par les LLM pour l'interrogation des données



SQL est un langage puissant et polyvalent utilisé pour gérer et manipuler les données stockées dans des bases de données. SQL a pour lui sa simplicité et sa facilité d'utilisation. Il utilise une syntaxe limitée qui permet aux utilisateurs d'écrire rapidement et facilement des requêtes pour extraire, modifier et manipuler les données dans la base de données. Cela en a fait le langage privilégié pour les personnes qui doivent accéder aux données de leur organisation et les analyser.

SQL fournit un langage commun qui peut être utilisé pour accéder aux données et les analyser à partir de diverses sources, notamment les feuilles de calcul, les fichiers plats, les bases de données NoSQL et les entrepôts de données traditionnels. Avec la prolifération des plateformes de données et la quantité croissante de données générées, l'utilisation du SQL est prépondérante chez les Data Analyst, Data Engineer et Data Scientist mais aussi certains profils plus "low-code" comme des business analysts ou des Citizen Developers.

Vers la généralisation des requêtes en langage naturel

Si SQL reste un langage universel, encore largement utilisé et mis à jour, les avancées récentes allant jusqu'à l'intégration de modèles d'IA et de ML directement dans le langage, ces mêmes technologies pourraient progressivement être amenées à le remplacer.

Une des promesses des LLM et de l'IA générative se trouve dans la possibilité pour l'utilisateur d'interagir en langage naturel directement avec un système d'information (le LLM traduisant l'interrogation en requête SQL). Il n'y a donc qu'un pas à franchir pour envisager de questionner directement ses bases de données : "Quels sont les résultats du Q3 ?" ou encore "Liste nos 5 produits les plus importants en terme de CA".

Les géants de la tech comme Google ou Microsoft s'attaquent déjà à ce marché de l'analytique et de la business intelligence, et les premières démonstrations de leurs assistants d'interrogation en langage naturel esquissent déjà les contours d'un futur sans SQL à manipuler - sauf par les experts du data engineering.

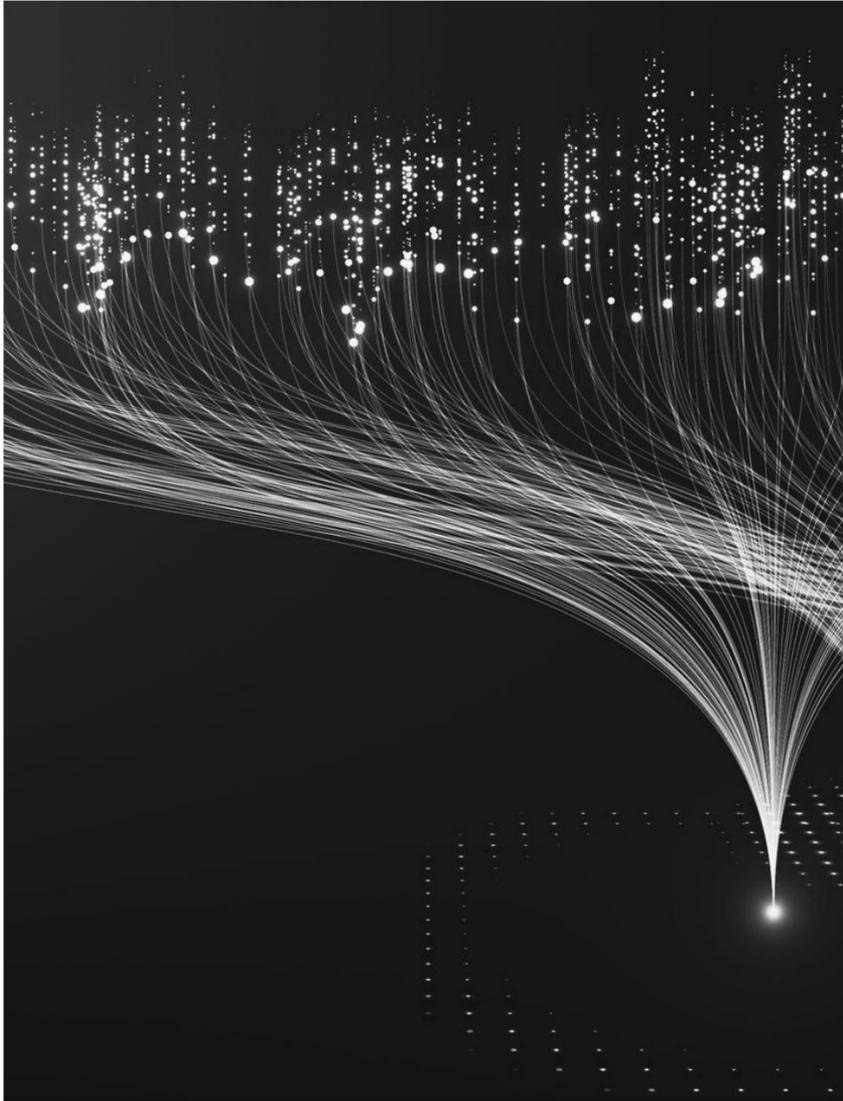


DATA TECH
#Data Valorisation

Comment utiliser la donnée

#Data #DataConcepts #ETLvsELT

L'ELT détrône l'ETL



L'extraction des données sources et leur chargement dans les systèmes cibles a longtemps été l'apanage des ETL (Extract, Transform & Load). A l'époque où il s'agissait de capter les données d'un système transactionnel pour les injecter au bon format au sein des entrepôts de données (datawarehouses), les outils ETL réalisaient toutes les opérations de transformation nécessaires entre les deux.

L'avènement des nouvelles architectures de données privilégie le plus souvent le chargement des données brutes au sein d'un datalake. L'étape de transformation est réalisée ensuite, si elle s'avère nécessaire, pour injecter les données au sein du datawarehouse. De cette façon, les data scientists auront accès aux données brutes et, si de nouveaux besoins analytiques émergent, de nouvelles transformations pourront être opérées à partir des données brutes.

Ce changement vers l'ELT (extraction, chargement, puis transformation) a aussi été rendu possible par les nouvelles capacités offertes par les plateformes analytiques Cloud, où l'usage du processeur et de la mémoire n'est plus aussi contraint que dans les datawarehouses traditionnels.

Cela bouleverse forcément le marché des outils d'ingestion de données, qui a vu aussi naître des outils qui se consacrent spécifiquement à la transformation. Le premier d'entre eux est le framework dbt, qui jouit d'une énorme popularité - notamment parce qu'il permet de décrire les transformations de données de façon modulaire, de les tester et de les documenter - la documentation produite intégrant automatiquement le lignage de la donnée. Du point de vue du développeur, cependant, la qualité du code peut laisser à désirer, ce qui conduit des entreprises à s'intéresser à un framework plus récent, Dataform (racheté puis intégré à Google Cloud Platform), qui doit encore améliorer ses fonctionnalités.

Ce découpage entre EL & T et l'évolution vers la Data-as-code nous paraissent en tout cas actés.

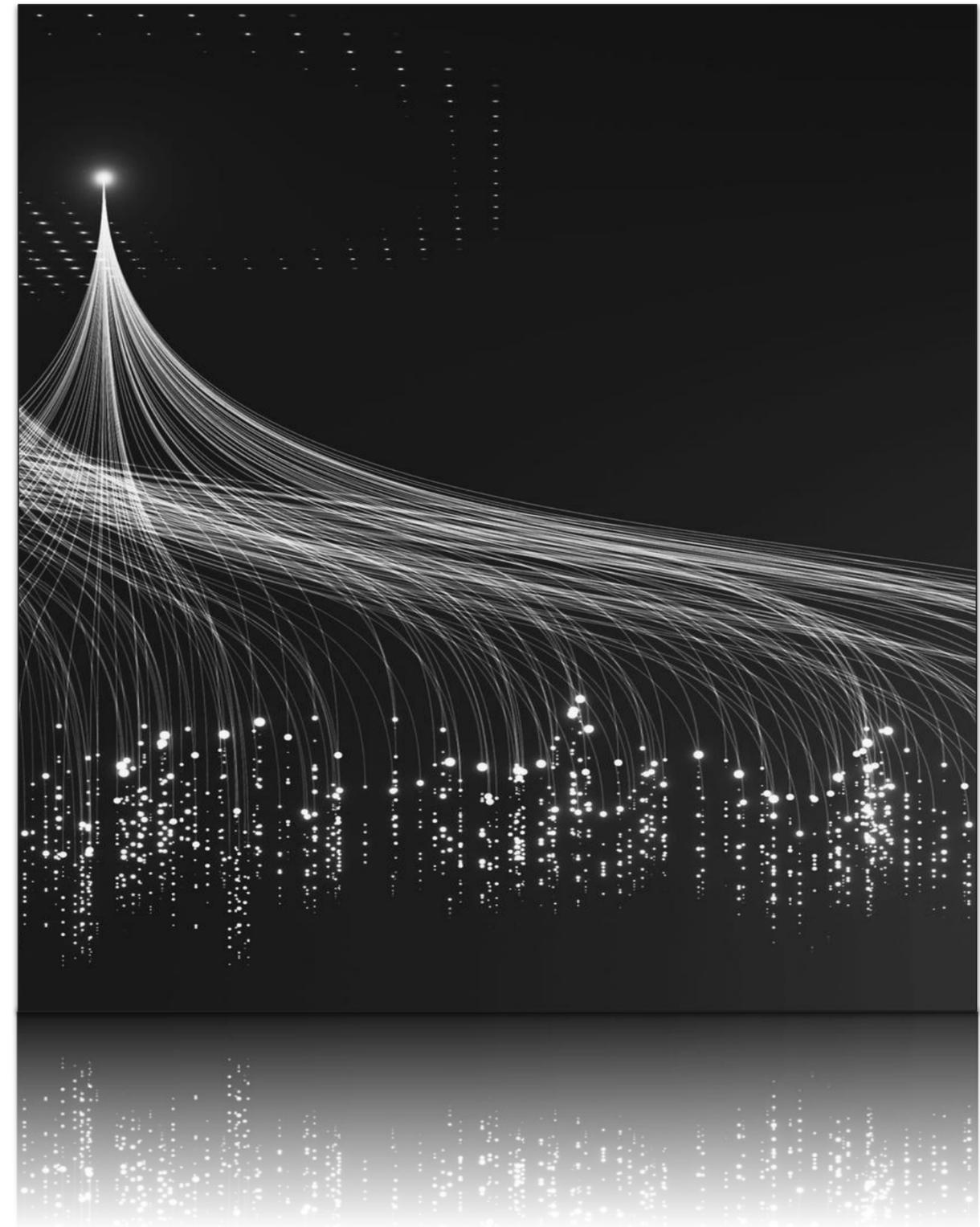
#Data #DataConcepts #StreamingAnalytics

Streaming Analytic

L'analyse de données en temps réel, également appelée "streaming analytics", est devenue incontournable pour les entreprises cherchant à prendre des décisions éclairées et rapides. Initialement utilisée pour les uses cases nécessitant une forte réactivité comme la détection de fraude ou la surveillance d'activités en temps réel dans l'industrie, cette technologie s'étend désormais au domaine de la Business Intelligence et du reporting.

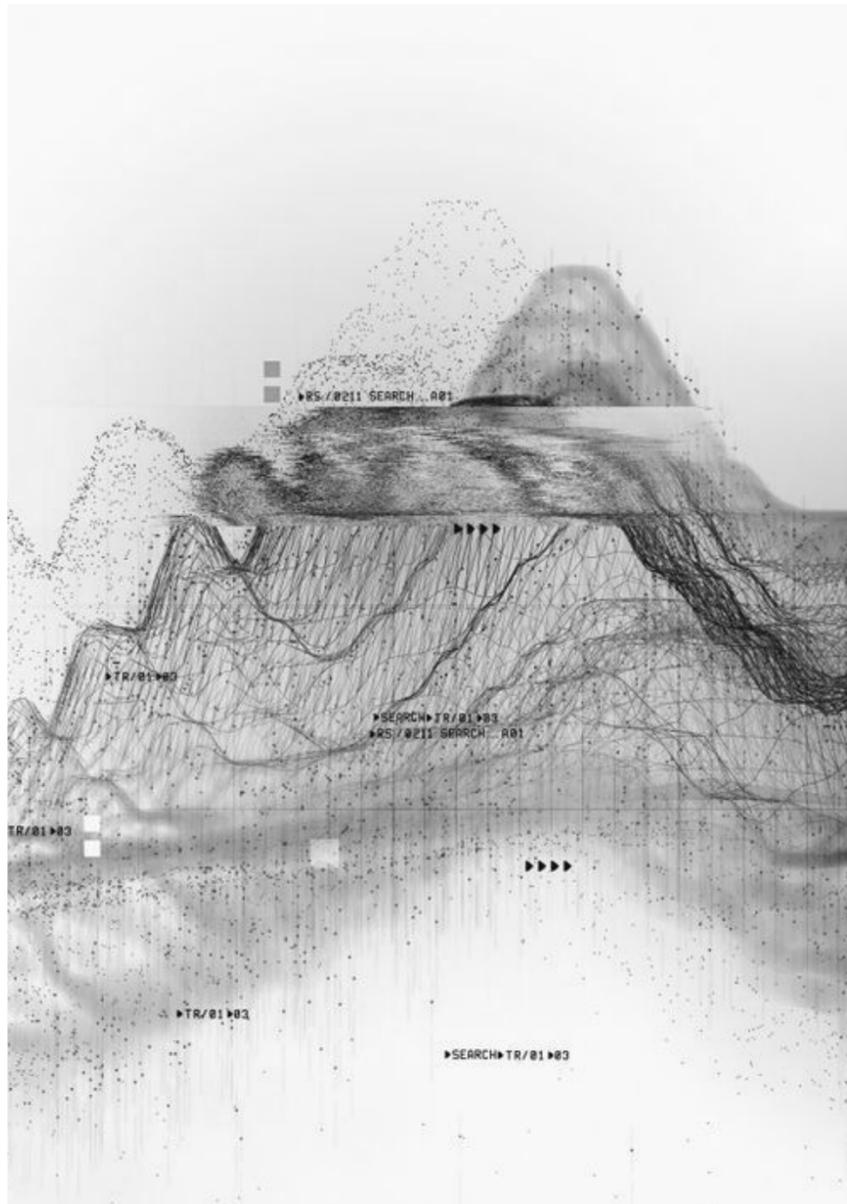
Pouvoir observer en direct l'impact d'une campagne marketing ou d'une promotion sur un site e-commerce permet aux utilisateurs business de mieux piloter leur activité. De plus, avec l'essor des agents conversationnels et des modèles de langage (LLMs), **l'accès à des données en temps réel améliore la précision des réponses fournies aux utilisateurs.**

Cependant, la volatilité de l'information dans de tels architecture implique la nécessité d'une stratégie de streaming analytics solide. Maîtriser les flux de données et **mettre en place une gouvernance efficace** sont essentiels. Les équipes informatiques et métiers doivent collaborer pour assurer la fiabilité des données, la gestion des données personnelles et le traitement des erreurs afin d'avoir une chaîne de donnée fiable et pertinente.



#Data #DataConcepts #ETLvsELT

Business intelligence & Data Visualization



La **Business Intelligence** (BI) et la **Data Visualization** jouent un rôle crucial dans les organisations en facilitant la prise de décisions finales. L'émergence de l'approche "as code" apporte une nouveauté intéressante en intégrant des **contrôles de non-régression** dans les pipelines de données, assurant ainsi la **qualité des données malgré les mises à jour fréquentes**. Cette tendance souligne une évolution significative dans la manière dont les données sont gérées et utilisées.

Parallèlement, nous observons une convergence croissante entre la BI et la data. **Autrefois considérées comme des univers distincts**, ces deux disciplines se rapprochent désormais grâce à des **couches sémantiques** qui centralisent la **gestion des accès**, les **politiques de sécurité**, et le **calcul des KPIs**. Cette intégration renforce la cohérence et la fiabilité des données, tout en favorisant une collaboration plus étroite entre les équipes chargées de la gestion de la data et celles responsables de la BI.

En résumé, la BI évolue vers une approche plus **intégrée** et **automatisée**, renforçant ainsi son rôle stratégique au sein des organisations. La convergence avec la data et l'adoption de l'approche "as code" ouvrent de nouvelles perspectives pour une prise de décision plus éclairée et rapide, tout en garantissant la qualité et la cohérence des données sur le long terme.

DATA ORGANISATION

LES CONTRIBUTEURS
DE CETTE SECTION

ILS ET ELLES ONT ÉCRIT LES TECH TRENDS 2024



Sandrine BOITEAU
Consulting Director Orga & Product,
WENVISION



Aude DEFRETIERE
Consultante Product Strategy,
WENVISION



Romain MAZUIR
Head of Product Strategy,
WENVISION



Vanessa PERILLAT
Group Chief Talent Officer,
SFEIR | WENVISION



Olivier RAFAL
Consulting Director Strategy,
WENVISION



Florent LEGRAS
CTO Data Groupe,
SFEIR



DATA ORGANISATION
#DataConcepts

Data Mesh, pour un usage croissant et décentralisé



Le Datamesh est un cadre conceptuel qui englobe plusieurs dimensions et repose sur la notion de fédération. Nous connaissons ses aspects d'architecture du SI : découper les systèmes informatiques en domaines indépendants et autonomes associés à une approche produit, les doter d'une couche d'interface (API ou autre) et de la documentation associée pour que le reste de l'organisation puisse venir consommer les produits.

En 2023, le datamesh s'est particulièrement illustré dans ses impacts organisationnels en entreprise. L'article fondateur de Zahmak Deghani de 2019 propose d'ailleurs avant tout un moyen de sortir de l'impasse aux entreprises qui ont une organisation complexe, morcelée et parfois sillotée. C'est un chemin possible pour ces entreprises, qui n'ont pas réussi à créer un point de vérité central d'entreprise avec une gouvernance forte et un partage exhaustif du patrimoine de données via un datalake maîtrisé et performant.

Le Data Mesh impose une approche Produit, avec une gouvernance appropriée

D'un point de vue organisationnel, le datamesh propose de s'aligner sur la vision Produit, à savoir : chaque unité de production (souvent un domaine data croisé d'un domaine métier) est responsable de bout en bout des produits qu'elle est en capacité de proposer à ses clients. L'unité de production dispose de tous les moyens nécessaires pour produire et est en capacité de prendre des décisions sur les évolutions à implémenter pour servir au mieux ses clients.

La gestion de la plateforme technique reste la prérogative de la DSI, qui elle aussi adopte le mode d'interaction du produit et travaille à mettre à disposition son produit : la plateforme data composée de l'ensemble des éléments de stockage, des flux et outils d'exploitation et surveillance associés au périmètre data.

Deux éléments clés d'un déploiement sans couture d'une organisation data mesh :

- la **data gouvernance** avec des règles et standards uniformes pour l'ensemble de l'entreprise. La data gouvernance peut être traitée par une équipe "chapeau" de l'organisation data si elle a ses points de relais dans les différents domaines. Sinon elle peut être intégrée au dessus de l'équipe plateforme et s'associer aux équipes plateforme.
- la **gestion du patrimoine data** de l'entreprise qui doit être matérialisé dans un set documentaire partagé avec l'ensemble des collaborateurs. La documentation des données, de leurs transformations et des domaines owners est soit décentralisée dans les domaines, soit centralisée au dessus de l'équipe plateforme. La donnée est alors pensée "as-a-service" et constitue une brique complémentaire du mesh.

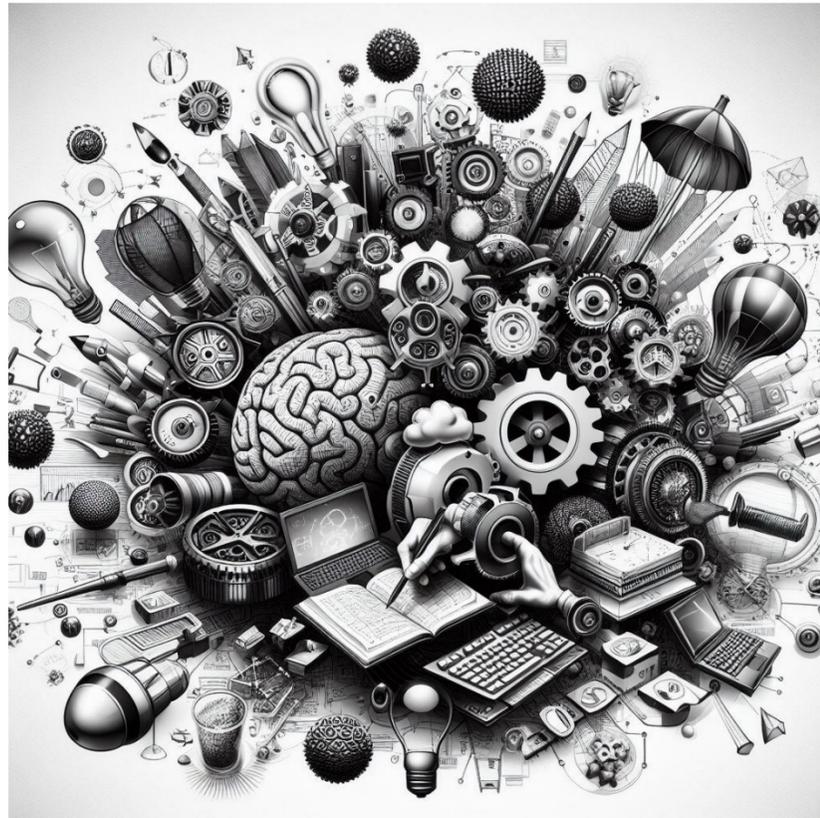
Le datamesh c'est donc un peu du meilleur des mondes adapté aux enjeux spécifiques de la data pour les entreprises. Le framework est utile et a déjà prouvé son efficacité chez un grand nombre de clients, mais on ne peut faire l'économie d'une gouvernance des données maîtrisée et déployée à l'échelle au risque d'aboutir à des bastions et de retomber dans les problèmes d'enclavement des données.

DATA ORGANISATION

#Gouvernance

#Orga #Equipes #dataproduit #engineering

Data Product Engineering : bien intégrer la data avec le SI



Comment délivrer des produits Data intégrés? Nous pensons que le concept de Data Product Engineering doit émerger : une organisation Data qui utilise les mêmes leviers d'ingénierie qu'une organisation Produit.

Il n'y a aucun doute sur le fait que la data soit devenue un actif stratégique pour les entreprises. Elle est utilisée pour prendre des décisions, améliorer les produits et services, et créer de nouveaux modèles économiques.

Ce qui est moins évident, c'est que la data est un produit comme un autre. Elle doit être collectée, traitée et valorisée de manière efficace et efficiente. Et elle dépend aussi de son stockage, des capacités du réseau, de l'architecture, de normes réglementaires et de sécurité. Comme tout produit, sa production peut être industrialisée, au sein d'une Factory.

La complexité de ce produit, c'est qu'il nécessite une étroite collaboration entre des équipes data (qui connaissent la donnée et ses traitements), des responsables produits (finalité des produits data, cycle de vie, pilotage par la valeur...) et des équipes IT (plateforme et pipelines de données, outils de consultation...).

Chaque entreprise doit mettre en place un TOM adapté à ses caractéristiques

Comment assurer une collaboration étroite entre ces équipes et profils ? Malheureusement, il n'existe pas de méthodologie unique pour la data et la fluidification depuis le besoin jusqu'à l'usage en production est essentielle pour la réussite des produits data. Il revient à chaque entreprise de construire son modèle opérationnel (TOM, target operating model), adapté à son organisation et sa culture. Il existe tout de même des patterns, qui pourront être employés pour accélérer la configuration de ce TOM et son déploiement, et s'assurer de l'application de bonnes pratiques.

Dans tous les cas, nous constatons qu'après les initiatives de DataLab isolés, les Data Factories font face à leur dépendance avec le reste du SI - et c'est un signe de maturité. Dans ce contexte, les entreprises vont devoir **adopter une approche d'ingénierie de la data, s'appuyant sur des processus et outils de livraison bout en bout, ainsi que des principes inhérents aux organisations orientées produits, comme les approches Lean et Agile.**



#Data #DataConcepts #DataCatalogs

Le data cataloging à l'heure de la GenAI



Avant, les data catalogs n'étaient jamais à jour. Trop fastidieux à remplir. Demain, les métadonnées seront automatiquement mises à jour. C'est l'un des plus beaux cas d'usage de l'IA générative : révolutionner le data management en s'appuyant sur les LLM. Un changement de paradigme aussi important que le passage de l'annuaire Yahoo au moteur de recherche Google.

Le catalogues de données sont au cœur du data management. Un data catalog fournit une vue centralisée et consolidée des données d'une entreprise. Il permet de faciliter la découverte, la compréhension et l'utilisation des données par les collaborateurs de l'entreprise. Plus précisément, un data catalog permet de :

- Décrire les données : source, format, type, définition, qualité, etc. Ces informations permettent aux utilisateurs de comprendre les données et de les utiliser de manière appropriée.
- Répertorier les données : par sources, propriétaires, utilisations, etc. Cela permet aux utilisateurs de trouver les données dont ils ont besoin.
- Rendre les données accessibles en fournissant des outils de recherche et de découverte. Cela facilite l'utilisation des données par les utilisateurs.

Le problème, c'est que pour rendre ces services, les data catalogs doivent être alimentés et mis à jour manuellement. Un processus laborieux qui ne peut pas suivre le rythme exponentiel de la croissance des données. Ce modèle est donc synonyme d'inefficacité et un fardeau coûteux pour les entreprises, qui induit une certaine défiance vis-à-vis du data cataloging..

Aujourd'hui : l'inspiration Google

L'automation change la donne. Similairement à l'idéologie de Google, un data catalogue piloté par IA mise sur l'automatisation pour crawler, indexer et référencer les données, boostant ainsi la pertinence et la facilité d'accès. Avec leur capacité de traitement du langage naturel, les LLM offrent des solutions pour traduire des volumes massifs de données en informations pertinentes et exploitables, simplifiant considérablement les tâches de recherche et d'analyse pour les utilisateurs à tous les niveaux de l'entreprise.

Face à la complexité croissante et à la quantité de données générées, l'IA générative est non seulement une aubaine mais une nécessité pour réinventer les méthodologies de data management. La GenAI devrait donc devenir le pilier d'un système de navigation des catalogues de données d'entreprise où trouver l'information désirée sera aussi simple que de poser une question à un moteur de recherche grand public.

#Produits #ProductTrends #DataProducts #AIProducts

Une approche produit aussi pour la Data et l'IA



L'approche produit consiste à considérer les données comme un produit qui doit être conçu, développé et commercialisé. Cela implique de mettre l'accent sur les besoins des clients et des utilisateurs, et de travailler en collaboration avec les équipes commerciales pour identifier les opportunités de valorisation des données. Développer une approche produit au sein du pôle Data agit ainsi comme un catalyseur permettant de transformer la donnée en un levier stratégique.

Cette approche produit de la data amène 3 bénéfices clés :

- Une meilleure compréhension des besoins clients. L'approche produit conduit l'équipe data à se concentrer sur les besoins des clients et des utilisateurs, afin de délivrer des produits et des fonctionnalités qui répondent réellement à leurs attentes.
- Une plus grande agilité et réactivité. Dans l'approche produit, l'équipe data est en relation plus étroite avec le métier, et donc mieux à même d'identifier les problèmes métiers sous-jacents, permettant ainsi une réflexion proactive pour fournir des solutions axées sur les besoins concrets du business.
- De nouvelles perspectives pour la valorisation des données. En adoptant une vision centrée sur le client, l'équipe data peut développer des produits et des fonctionnalités qui enrichissent considérablement les services ou produits de l'entreprise, en créant de nouveaux assets stratégiques ou en exploitant de nouvelles niches de marché.

La mise en place d'une approche produit au sein d'une équipe data suppose que la culture soit adaptée : adopter l'orientation produit est une démarche culturelle avant d'être méthodologique. L'équipe data devra soit travailler en étroite collaboration avec une équipe produits, amenant les rôles et les processus adaptés, soit mettre en place ces rôles et processus en interne. Dans tous les cas, des mécanismes de mesure et de suivi devront aussi être mis en œuvre.

#Data #DataGouv #DRHData

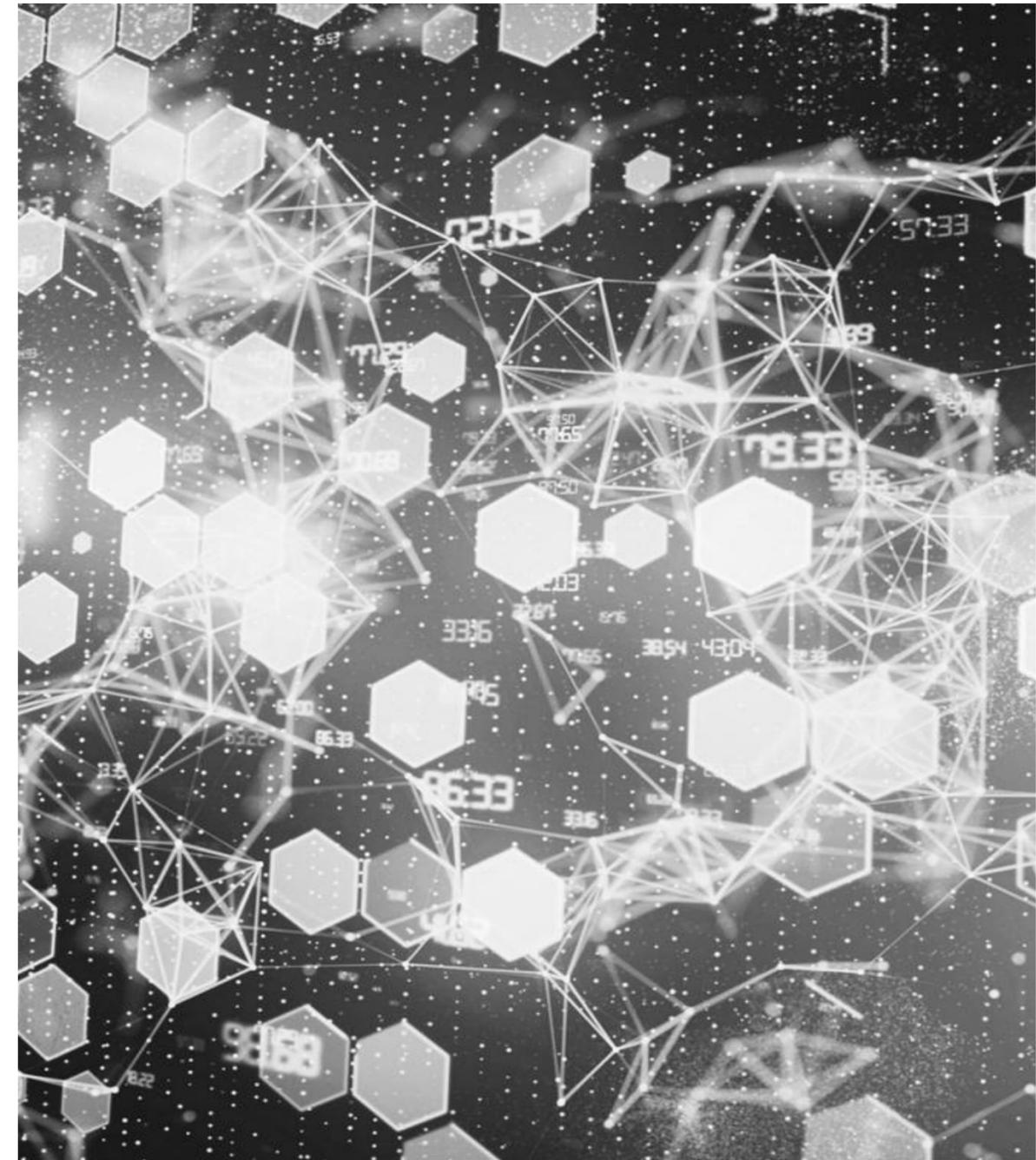
Des DRH Data pour catalyser les organisations Data

Dans le paysage data-driven de 2024, un rôle majeur doit prendre enfin toute sa place : celui du Directeur des Ressources Humaines Data. Son implication devient incontournable pour piloter la stratégie talent au cœur des organisations. En collaboration avec le CDO et les équipes Data, il est le catalyseur essentiel pour fédérer les compétences et acculturer les collaborateurs à la valeur stratégique des données, un atout indispensable pour les entreprises qui souhaitent pérenniser leur organisation.

Les rôles essentiels d'un ou une DRH Data

- **L'intégration des talents.** Le DRH Data endosse le rôle crucial d'intégrer les talents essentiels aux besoins de l'organisation. Sa mission commence par une cartographie détaillée des postes et responsabilités, fondamentale pour le déploiement d'une gouvernance des données pérenne.
- **La formation des talents.** Conscient de l'évolution constante des technologies, il met en œuvre des programmes de formation personnalisés. Il s'assure que chaque membre de l'organisation, quels que soient son rôle et son niveau, accède aux formations nécessaires pour s'adapter et exceller dans son domaine. Ces initiatives soutiennent et stimulent le déploiement d'une culture data globale.
- **La reconnaissance des talents.** Le DRH Data veille à instaurer une culture du feedback où la reconnaissance est centrale. Signe distinctif d'une culture d'entreprise progressiste, il valorise les réalisations individuelles et collectives et offre des possibilités d'évolution à travers un parcours professionnel riche et data-orienté. La reconnaissance motive les talents et stimule un sentiment d'appartenance à l'entreprise.

Investir dans le rôle du DRH Data est une étape décisive pour toute entreprise ambitieuse. C'est lui qui va piloter l'asset le plus précieux des organisation data driven : le capital humain.



La gouvernance de la donnée, gardienne de la fiabilité des LLM



La montée en puissance de l'IA générative, illustrée par les avancées de modèles comme Gemini ou GPT-4, a marqué un tournant technologique majeur, prompt à transformer notre façon d'interagir avec les données. À l'horizon 2024, la gouvernance de l'IA s'impose comme la pierre angulaire permettant de tirer pleinement parti de ces technologies de pointe, tout en atténuant leurs risques inhérents.

La gouvernance de l'IA est un investissement capital pour les entreprises souhaitant naviguer avec succès dans l'écosystème de l'IA en 2024. Elle est la garante d'une technologie IA générative sécurisée, conforme et éthique, assurant tant l'excellence opérationnelle que la valorisation des données.

A contrario, de nombreux cas concrets initiés en 2023 démontrent comment une gouvernance de données défaillante peut mener à des incohérences opérationnelles, telles que des recommandations de produits erronées, ou encore des accès à une donnée non autorisée. Ce constat souligne l'impératif d'une gouvernance adéquate pour l'intégration réussie de l'IA générative dans les processus métier.

Les grands principes à garder en tête

- **La qualité des données comme socle de la fiabilité**

La réussite des systèmes d'IA générative repose sur un adage intemporel : une qualité de données irréprochable conditionne l'exactitude des résultats. Une gouvernance de données robuste assure l'intégrité et la sécurité des datasets, garantissant ainsi fiabilité et pertinence des LLM.

- **L'éthique pour une utilisation sécurisée et responsable**

La gouvernance de l'IA doit veiller à la conformité des données vis-à-vis des réglementations en vigueur, notamment le RGPD en Europe, et s'assurer du respect de la vie privée. Elle doit également prévenir les biais et veiller à l'équité, amplifiant le caractère éthique de l'utilisation de l'IA.

- **Traçabilité et transparence comme leviers**

En instaurant des politiques transparentes et une traçabilité des décisions, la gouvernance de l'IA favorise une responsabilité claire et répond aux exigences réglementaires. Un tel cadre renforce la confiance des utilisateurs et des parties prenantes.

#Data #DataConcepts #DataQuality

Data Quality



La qualité des données représente un défi majeur pour toute entreprise. Souvent, **l'urgence de déployer rapidement des pipelines pour générer de la valeur et réduire le time to market** relègue la qualité des données au second plan, entraînant des **écarts entre les indicateurs** et une **absence de source unique de vérité**.

Pour remédier à cela, il est crucial d'intégrer des **processus de validation** et de **nettoyage automatiques** dans les pipelines de données. Cela comprend la **détection des doublons**, la **correction des valeurs aberrantes**, la **normalisation des formats**, et la **validation des contraintes de données**. De plus, l'adoption de pratiques de gouvernance des données, telles que la définition de **normes de qualité** et la **documentation des processus de nettoyage**, est essentielle pour maintenir la qualité des données sur le long terme.

L'automatisation des tests de qualité des données, intégrés aux pipelines et exécutés régulièrement, permet de détecter rapidement les problèmes et d'appliquer des correctifs appropriés, garantissant ainsi la fiabilité des données à travers tout le système d'information.

Enfin, l'utilisation de **métriques de qualité des données**, telles que le **taux d'exactitude**, le **taux de complétude** et le **taux de conformité**, permet d'évaluer objectivement la qualité des données et d'identifier les domaines nécessitant des améliorations continues, contribuant ainsi à maintenir la cohérence des données dans toute l'entreprise.

#Data #DataGouv #DataContract

Data Contract

Les data contracts, émergents depuis quelque temps, représentent un ensemble de **directives** et **d'engagements** définis pour **assurer la qualité**, la **cohérence** et la **gouvernance** des données au sein d'une organisation. Ils établissent un **cadre formel** qui attribue clairement **l'ownership** des données à des entités spécifiques, telles que les producteurs et les consommateurs de données.

Au cœur de ce concept réside la notion de responsabilité partagée et de transparence dans la gestion des données. Par exemple, un producteur de données se voit assigner un data contract détaillant les spécifications du fichier plat associé, ainsi que le schéma attendu de celui-ci. De cette manière, il est clair pour toutes les parties concernées quelles sont les attentes en termes de qualité et de structure des données.

De plus, les data contracts sont souvent associés à des mécanismes de communication intégrés dans les pipelines CI/CD. Cette approche permet une réactivité accrue face aux changements de schéma ou de format de données, en automatisant la notification des propriétaires concernés en cas d'incohérence détectée.

En résumé, les data contracts représentent un outil essentiel pour instaurer une gouvernance des données efficace, en garantissant une **responsabilité claire** et une communication transparente tout au long du cycle de vie des données au sein de l'organisation.



#Data #DataConcepts #DataLineage

Le Data Lineage : un outil indispensable pour la fiabilité



Dans le paysage actuel des systèmes informatiques, la complexité est omniprésente. Les données transitent par une multitude d'applications, subissent de nombreuses transformations et alimentent divers processus. Cette complexité engendre un manque de visibilité sur l'origine, le parcours et l'utilisation des données, ce qui peut entraîner des erreurs, **une perte de confiance et une inefficacité opérationnelle**.

Une solution existe pour remédier à ce problème : le Data Lineage. **Cette discipline consiste à cartographier le cycle de vie complet d'une donnée, de sa source à sa destination finale**, en passant par toutes les étapes de transformation. Le Data Lineage permet de répondre à des questions cruciales telles que :

D'où provient cette donnée ?

Quelles transformations a-t-elle subies ?

Où est-elle utilisée ?

Qui est responsable de sa gestion ?

En adoptant une approche de Data Lineage, les organisations peuvent bénéficier de nombreux avantages. En comprenant l'origine et le parcours des données, il est plus facile d'identifier et de corriger les erreurs, ce qui **renforce la confiance dans les analyses et les prises de décision**. En plus de cela elle permet de rationaliser les flux de données, d'éliminer les redondances et d'identifier les goulots d'étranglement, ce qui améliore l'efficacité opérationnelle.

Elle facilite la mise en conformité avec les réglementations en matière de protection des données, telles que le RGPD, en permettant de tracer l'utilisation des données personnelles.

Le Data Lineage est un **outil indispensable pour toute organisation souhaitant améliorer la fiabilité, l'efficacité et la conformité de ses données**. En cartographiant le cycle de vie des données, le Data Lineage permet de prendre des décisions éclairées, d'optimiser les processus et de renforcer la confiance dans les informations utilisées.

#Tech #GenAI #Challenges

La Data Observability : votre alliée pour une donnée de qualité



Dans un monde de plus en plus data-driven, la fiabilité et la qualité des données sont cruciales pour la prise de décision et le bon fonctionnement des opérations. Pourtant, les entreprises font face à des défis croissants pour maintenir une vision claire et en temps réel de l'état de leurs données. Des erreurs dans les pipelines de traitement, des changements inattendus dans les schémas de données ou des dérives progressives des valeurs peuvent passer inaperçus, **impactant négativement les analyses et les modèles de machine learning.**

La Data Observability émerge comme une solution essentielle pour répondre à ces défis. Il s'agit d'une discipline qui permet de surveiller en temps réel l'état de santé des données et des pipelines de traitement. En collectant et en analysant des métriques clés, la Data Observability offre une visibilité complète sur les données, permettant de :

Détecter les anomalies et les erreurs : Identifiez rapidement les problèmes tels que les valeurs manquantes, les formats incorrects, les duplications ou les incohérences dans les données.

Surveiller les performances des pipelines : Suivez les temps d'exécution, les volumes de données traités et les taux d'erreur pour identifier les goulots d'étranglement et optimiser les performances.

Analyser les dérives de données : Détectez les changements subtils dans la distribution des données, les valeurs aberrantes ou les tendances inattendues qui pourraient indiquer des problèmes de qualité.

Améliorer la qualité des données : En comprenant l'origine et l'impact des problèmes de données, des mesures correctives peuvent être prises pour améliorer la qualité et la fiabilité des données.

En adoptant la Data Observability, vous investissez dans la fiabilité et la qualité de vos données, éléments essentiels pour la réussite de vos projets data-driven.



/03

IA

IA

LES CONTRIBUTEURS
DE CETTE SECTION

ILS ET ELLES ONT ÉCRIT LES TECH TRENDS 2024



Ravidhu DISSANAYAKE

Consultant Data / IA,
WENVISION



Martin ELIARD

Head of Data Governance,
WENVISION



Marie FONTAINE

Head of AI,
WENVISION



Florent LEGRAS

CTO Data Groupe,
SFEIR



Mylène MAIGNANT

Consultante IA,
WENVISION



Aurélien PELLETIER

Consulting Director Platform,
WENVISION



Baptiste PUGNAIRE

Consultant Data / IA,
WENVISION



Olivier RAFAL

Consulting Director Strategy,
WENVISION



Céline THOORIS

Consulting Director Data,
WENVISION

IA

#Predictive Analysis

#Tech #GenAI #ChatUI

L'IA révolutionne nos interactions avec le SI



L'interaction avec un SI devrait être aussi simple que de converser avec un collègue. Et pourtant, depuis l'arrivée des "menus" dans les applications, nous ne cessons de cacher l'information en la rendant accessible uniquement après une suite d'interactions. L'objectif d'une interface conversationnelle est de simplifier et de rendre plus intuitive l'accès aux données.

Il n'y a qu'à observer le succès quasi-instantané de ChatGPT pour comprendre le potentiel de ce mode d'interaction. Alors que la technologie existait depuis plusieurs années, l'ouverture d'un accès en mode conversationnel a fait exploser son usage.

Les CUI (conversational user interfaces) ne sont pas seulement intuitives par nature, avec un besoin en formation minimale, elles éliminent également le besoin de naviguer à travers des menus hiérarchiques complexes et chronophages. En s'adaptant au niveau d'expertise de l'utilisateur, les CUI offrent une expérience personnalisée. L'expert peut poser ses questions d'expert, l'utilisateur occasionnel peut lui aussi interagir en posant ses propres questions.

Des CUI (conversational user interfaces) en complément des traditionnelles GUI

C'est la technologie qui se plie à la volonté de l'utilisateur : elle comprend le langage "parlé", et transforme la façon d'interagir avec les données, l'analytique et le prédictif.

L'avantage est l'accélération des tâches routinières par l'action de simples commandes. Que l'on consulte des bases de données, génère des rapports ou analyse des tendances, les CUI offrent une efficacité que les GUI (interfaces graphiques, graphical user interfaces) traditionnelles ne peuvent égaler.

Toutefois, il ne faut pas enterrer trop vite nos bonnes vieilles GUI ; elles ont toujours leur place. Pour les tâches complexes et spécifiques au sein du SI, la précision de la GUI classique reste. Ce parfait équilibre assure que, tandis que les CUI gèreront la majorité des interactions en améliorant la productivité, les GUI resteront disponibles pour gérer les interactions plus fines.

On n'a pas encore fait le tour du Machine Learning "traditionnel"



Nous avons assisté à l'émergence spectaculaire des Large Language Models (LLM) et de l'IA générative, qui bousculent les perspectives en offrant de nouvelles opportunités : génération de texte, synthèse vocale et créativité assistée. Cependant, elles ne sauraient totalement supplanter les méthodes classiques de ML pour plusieurs raisons.

Premièrement, tous les problèmes ne peuvent être traités avec succès à l'aide de l'IA générative et des LLM. Ces modèles peinent à résoudre des problèmes complexes, à inférer des lois physiques ou à s'adapter rapidement à de nouvelles situations. Deuxièmement, ce ne sont pas des systèmes déterministes : les réponses peuvent varier d'une fois à l'autre. Ensuite, les LLM reposent sur des architectures massives qui sont connues pour leur voracité énergétique et leur empreinte carbone substantielle, ce qui constitue un frein à leur adoption, en particulier dans les segments de marché où la durabilité et l'impact écologique constituent des facteurs de différenciation. Enfin, il est beaucoup plus difficile d'entraîner efficacement un modèle génératif sur ses propres données ce qui limite la personnalisation et l'adaptation à un contexte métier précis.

Le Machine Learning garde toute sa pertinence pour un large éventail de cas d'usage

Compte tenu de ces limitations, les techniques traditionnelles de ML conservent toute leur pertinence et leur valeur ajoutée dans les entreprises, en particulier on peut citer :

- Régression ou classification linéaire : dans les scénarios où les relations entre variables sont relativement simples, stables et linéaires, les méthodes classiques constituent des options privilégiées, en raison de leur facilité d'implémentation, de leur interprétabilité et de leur efficacité.
- Arbres de décision et forêts aléatoires restent des choix populaires dans les applications de diagnostic médical, de scoring de crédit, de segmentation marketing et de maintenance prédictive grâce à leur capacité à représenter des hiérarchies de décisions complexes et leur forte interprétabilité.
- Techniques de clustering, essentielles pour analyser et interpréter de gros volumes de données. Le clustering est fréquemment employé pour identifier et caractériser des segments de clientèle homogènes en fonction de leurs comportements d'achat ou encore pour anticiper les défaillances ou les anomalies en analysant des indicateurs de performance et de santé des machines, installations ou patients concernés.

D'autres applications du deep learning méritent également d'être mentionnées comme la Computer Vision ou le Reinforcement Learning. L'intérêt pour l'IA générative a permis à tous de redécouvrir les possibilités qu'offre le Machine Learning ; c'est l'opportunité pour les entreprises de repenser leur stratégie en y intégrant ces technologies innovantes sans céder aux "buzzword" du moment pour concevoir une stratégie IA cohérente et durable, intégrant ML et GenAI.



GEN AI

GEN AI

LES CONTRIBUTEURS
DE CETTE SECTION

ILS ET ELLES ONT ÉCRIT LES TECH TRENDS 2024



Ravidhu DISSANAYAKE
Consultant Data / IA,
WENVISION



Martin ELIARD
Head of Data Governance,
WENVISION



Marie FONTAINE
Head of AI,
WENVISION



Florent LEGRAS
CTO Data Groupe,
SFEIR



Mylène MAIGNANT
Consultante IA,
WENVISION



Aurélien PELLETIER
Consulting Director Platform,
WENVISION



Baptiste PUGNAIRE
Consultant Data / IA,
WENVISION



Martin ELIARD
Consultant Data / IA,
WENVISION



Céline THOORIS
Consulting Director Data,
WENVISION



Olivier RAFAL
Consulting Director Strategy,
WENVISION



Didier GIRARD
Co-CEO,
SFEIR & WENVISION



IA
#Generative AI



/04

CULTURE

L'IA générative est parmi nous - de manière durable

L'IA générative (Generative AI) est un type d'intelligence artificielle qui se concentre sur la génération de nouvelles données similaires à une entrée donnée. Des applications de génération d'images comme Dall-E 2, Stable Diffusion et Midjourney, ou de génération de texte ou de code comme GPT / ChatGPT ou encore GitHub CoPilot ou AWS Code Whisperer, ont très rapidement mis ces capacités sur le devant de la scène.

3 techniques clés

Une des techniques clés utilisées en IA générative est l'utilisation de réseaux génératifs adverses (generative adversarial networks / GAN), qui se composent de deux composants : un générateur et un discriminateur. Le générateur produit des données synthétiques basées sur l'entrée, tandis que le discriminateur tente de déterminer si les données sont réelles ou fausses. Cette compétition entre les deux réseaux permet à l'IA de générer des données de plus en plus similaires à l'entrée.

Les auto-encodeurs à variation (Variational autoencoders / VAE) sont une autre technique importante utilisée en IA générative. Ces algorithmes se composent de deux éléments : un encodeur, qui comprime les données d'entrée en une représentation à dimension inférieure, et un décodeur, qui prend cette représentation et génère de nouvelles données similaires à l'entrée.

Les transformateurs (transformers) sont une troisième technique importante en IA générative. Ces algorithmes utilisent un mécanisme dit d'attention pour mesurer et prédire l'importance des différentes parties des données d'entrée. Les transformateurs sont à la base des progrès faits récemment dans le domaine du traitement automatique du texte (Le T de BERT / GPT).

Une infinité d'applications et de cas d'usage

L'IA générative est un domaine en évolution rapide qui suscite à juste titre énormément d'intérêt en raison de ses applications potentielles dans des domaines tels que l'augmentation des données, la génération d'images et le traitement du langage naturel. Cette technologie a de nombreuses applications potentielles, typiquement :

- Juridique : l'IA générative peut générer rapidement des contrats et autres documents juridiques, ou identifier des modèles dans des affaires juridiques qui pourraient aider à prédire les résultats de futurs cas ou à minima à construire des dossiers.
- Média : l'IA générative peut aider à créer de nouvelles images, vidéos, présentations ou autres formes de contenu à partir d'une entrée donnée, comme une photographie ou une description écrite, où à la manière d'un artiste

célèbre (la banque d'images Shutterstock a déjà conclu un accord pour commercialiser des images issues de Dall-E 2).

- Éducation : l'IA générative peut créer des plans de leçon personnalisés pour les élèves, ou générer automatiquement des questions de quiz et d'autres matériaux pédagogiques.
- Développement de logiciels : l'IA générative sait automatiser la création de code (pour interroger une API par ex.), ou aider à identifier les potentiels bugs et autres problèmes dans les logiciels.
- Assurance : l'IA générative peut contribuer à générer rapidement des polices d'assurance, ou identifier des modèles dans les données de réclamation qui pourraient aider à prédire les risques futurs.
- Banque : l'IA générative peut automatiser la création de rapports financiers, ou aider à écrire des messages commerciaux personnalisés.

Ce ne sont que quelques exemples des nombreuses applications potentielles de l'IA générative dans différents domaines. On voit d'ailleurs fleurir de nombreux projets d'apps ou de start-up. À mesure que la technologie continue d'évoluer, il est probable que de nouvelles utilisations innovantes seront découvertes.

#Tech #GenAI #UseCases

Les tendances d'usage de l'IA Générative en 2024



Les usages principaux de l'IA Générative peuvent être organisés autour de 2 grandes catégories : **la recherche et la création de contenus**. Partant de là, prétendre pouvoir lister les cas d'usage les plus pertinents pour les entreprises en 2024 relève de l'oracle tant l'horizon semble ouvert et tant la "bonne" utilisation d'une technologie repose sur les priorités stratégiques de l'entreprise. On peut néanmoins noter que la tendance est plutôt aux usages internes, moins risqués, maximisant plutôt les gains de productivité (assistants généralistes, assistants verticaux, assistants de programmation, générateur de contenu pour le marketing, la communication, les centres de relation client).

Les capacités en entreprise pour répondre aux besoins sont limités, nous observons donc plutôt un besoin de gouverner ces usages avec deux questions centrales : comment organiser leur collecte et surtout comment prioriser parmi ce champ des possibles ? L'application des principes du Product Management, avec l'ajout de quelques spécificités, répond à ces questions.

Collecter les besoins par l'usage d'une solution d'assistant interne

S'il ne fallait choisir qu'un usage pour 2024, celui de l'assistant généraliste pour les collaborateurs semble le plus intéressant. Il permet de lever des freins de productivité pour de nombreuses fonctions et il peut surtout servir de première marche pour construire un socle technologique, des compétences et collecter les besoins des collaborateurs par l'usage concret qui en est fait. Cela permettra par la suite de verticaliser des besoins à travers des assistants spécialisés par exemple. C'est le chemin que plusieurs de nos clients ont entrepris en 2023.

La priorisation des usages doit suivre les fondamentaux du product management

En premier lieu, il est absolument nécessaire de valider la réponse à un besoin et de vérifier l'adéquation des hypothèses de gain avec les objectifs stratégiques de l'entreprise. Cette première étape doit également d'ores et déjà inclure une vérification de la conformité de la finalité de l'usage au regard de la réglementation, à savoir l'AI Act.

Un facteur de risque peut aussi être appliqué en fonction de l'acceptabilité d'une réponse fautive produite par le système. C'est notamment pour cette raison que **les systèmes d'IA générative mis directement à disposition des clients seront rarement priorisés en 2024, au profit d'usages tournés vers l'interne** avec des hypothèses de gain de productivité.

La complexité doit être mesurée, avec quelques spécificités dont il faut tenir compte : l'accessibilité de la donnée, la maturité des briques technologiques en jeu, le degré de sécurité attendu.

Un framework tel que RICE (Reach, Impact, Confidence, Effort) adapté à l'IA Générative peut s'avérer un outil de gouvernance très efficace.

#Tech #GenAI #Gouvernance

Des IA génératives d'entreprise, pour limiter risques et shadow AI

Interdiction, blocage, bannissement : la presse regorge d'exemples d'entreprises ayant interdit ChatGPT. Si Samsung, Apple, Deutsche Bank, Verizon ou Amazon l'ont fait, c'est qu'il y a une raison : éviter la fuite de données. Il faut considérer les IA génératives grand public comme des photocopieuses en libre accès, qui peuvent à tout moment ressortir vos documents, briefs stratégiques, etc. Elles doivent donc être interdites. Ce qui veut dire que l'entreprise doit fournir un service équivalent dans un cadre sécurisé.

D'une part, parce qu'il serait idiot de se priver du potentiel de l'IA générative, mais aussi de monitorer l'usage qui en est fait. D'autre part, parce que les utilisateurs en entreprise trouvent toujours un moyen de contourner les interdictions lorsqu'ils trouvent quelque chose de beaucoup plus pratique. En l'occurrence, si l'usage de ChatGPT et des autres applications d'IA générative a explosé aussi rapidement, c'est bien parce que ces outils apportent une aide énorme dans le quotidien. Ne pas offrir de solution, c'est s'exposer à l'émergence rapide d'un "shadow AI", susceptible de mettre l'entreprise à risque.

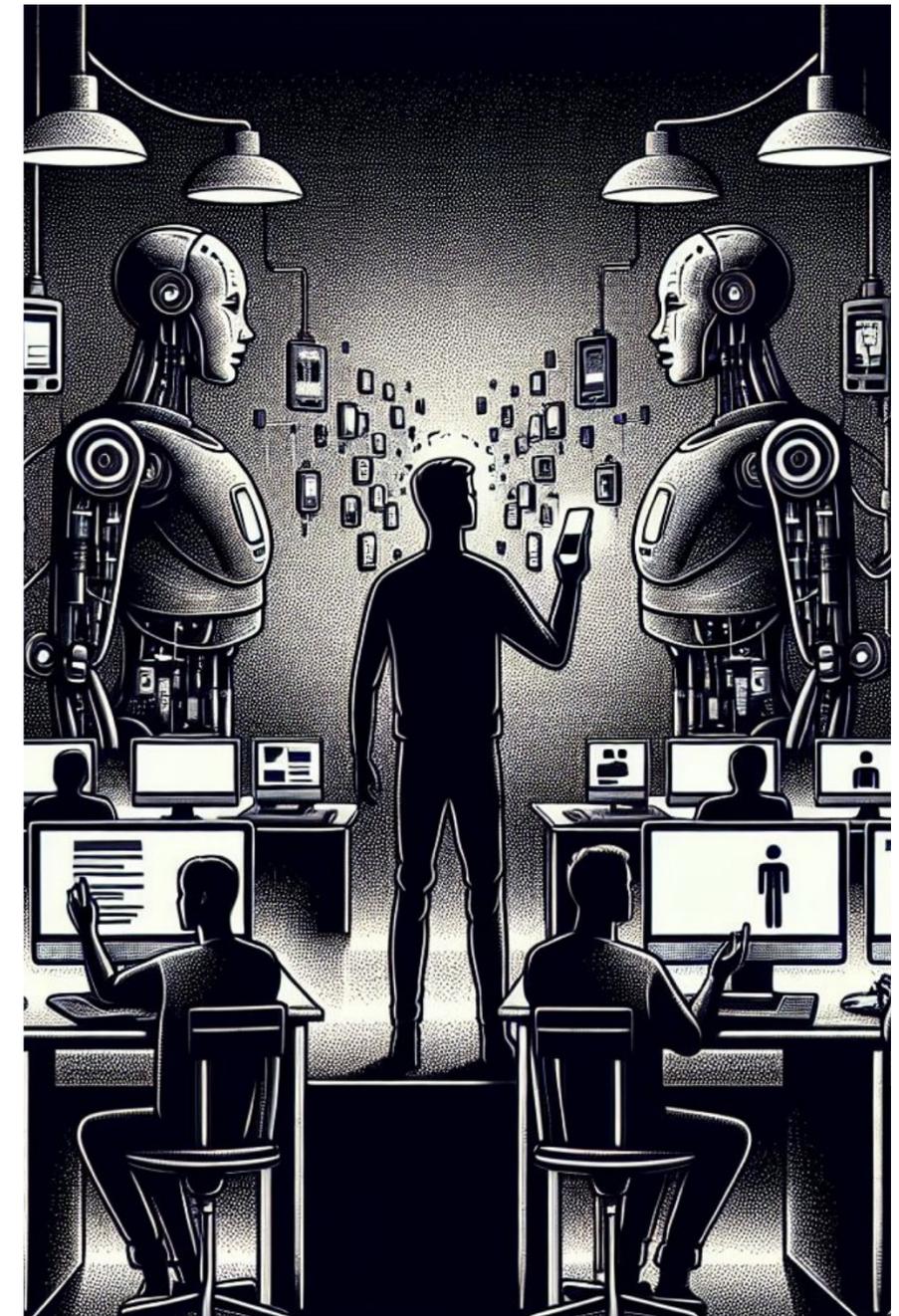
Une stratégie mixant "make" et "buy"

Plusieurs pistes sont envisageables pour mettre à disposition des environnements professionnels, sécurisés. Aucune ne consiste à développer son propre modèle. Il n'est généralement pas non plus pertinent de le "fine-tuner" pour optimiser son usage : c'est un processus lent et coûteux, à réserver à quelques cas précis.

Partir sur un environnement tout-en-un est possible mais risqué, car on s'en remet alors à un éditeur et sa roadmap, sans possibilité de supervision. L'approche pragmatique que nous préconisons et voyons plutôt se développer dans les entreprises consiste à mixer "make" et "buy". Autrement dit, s'appuyer sur les ressources des fournisseurs comme Google, Microsoft, Amazon, OpenAI, HuggingFace, Mistral, etc. et consacrer ses efforts à l'assemblage de ces ressources - par exemple une base de données vectorielle pour introduire de la sémantique au sein de ses propres documents, un orchestrateur type Langchain pour enchaîner les appels à des services ou encore des studios pour créer, optimiser et versionner les prompts utilisés.

Expliquer, acculturer

En plus d'interdire et de fournir une solution, il est impératif de communiquer et d'acculturer les collaborateurs à cette technologie et à ses enjeux. Il s'agit de faire comprendre pourquoi les outils grand public sont interdits, de montrer les forces ainsi que les faiblesses et les risques associés à la GenAI (biais, hallucinations, etc.), mais aussi de donner des perspectives sur les futures possibilités techniques qui seront mises à disposition et les cas d'usage qui pourront ainsi être mis en œuvre. Ces réflexions doivent s'inscrire dans un cadrage stratégique, qui lui-même doit respecter la vision et les valeurs de chaque entreprise.



#Tech #GenAI #Challenges

IA générative : 3 grands challenges sociaux en perspective



Les fake news

Alors que les technologies d'IA continuent de se développer à un rythme effréné, elles sont devenues à la fois une arme et un rempart dans la propagation et la détection des fausses informations. Le défi réside dans la création de modèles d'IA suffisamment intelligents pour discerner les faux des vrais faits (textes, images, vidéos...), tout en garantissant la protection de la liberté d'expression et en évitant les biais préjudiciables. Une gageure.

En 2024, les populations de 77 pays, dont les Etats-Unis, la Russie et l'Inde, seront amenées à voter. Dans ce contexte électoral mondial, la gestion des *faux* voire des *vrais faux* prend une dimension encore plus critique. Exemple frappant : au Pakistan, en décembre 2023, un parti politique a généré un discours de son leader, emprisonné, à partir de vidéos archivées, afin de pouvoir tout de même mener la campagne des législatives.

La consommation énergétique

La demande croissante en serveurs de données, en centres de calcul et en infrastructures

de refroidissement liée à l'IA générative contribue à l'augmentation de l'empreinte carbone du numérique. Les enjeux de la consommation d'énergie par l'IA en 2024 sont doubles : d'une part, il s'agit de réduire cette empreinte en développant des technologies et des infrastructures plus éco-énergétiques. D'autre part, il est primordial de sensibiliser et d'inciter les acteurs de l'industrie ainsi que les utilisateurs finaux à adopter des bons réflexes et des pratiques plus durables.

L'AI Act européen

L'AI Act en cours de discussion au sein de l'Union européenne illustre la prise de conscience croissante des enjeux liés à l'IA, notamment en matière de protection des droits fondamentaux et de prévention de la discrimination. Toutefois, sa future application soulève également des défis, notamment en ce qui concerne la mise en conformité des entreprises et l'équilibrage entre l'innovation et la réglementation. Même s'il est difficile de prévoir à quelle échéance ce texte sera en vigueur, il est clair que d'ores et déjà, la régulation de l'IA s'annonce comme un défi majeur à prendre en considération dès cette année.



/04

PLATEFORME

#Plateformes #Archi&Dev #GenAI

Les briques essentielles d'une plateforme GenAI



Une plateforme GenAI permet de répondre aux besoins en matière d'IA générative de la manière la plus efficace possible. Elle ouvre un monde de possibilités pour innover et améliorer les opérations de l'entreprise, en gardant ouvertes toutes les options possibles en matière de fournisseurs, de modèles, de cas d'usage ou encore d'intégration avec le SI et les processus de l'entreprise.

Les composantes a minima d'une plateforme GenAI :

- **Chat intelligent** : avoir votre propre version de ChatGPT grâce à l'accès à un ou des LLM associés à une interface de chat.
- **Bibliothèque de connaissances** : pouvoir enrichir les connaissances du LLM avec vos documents, vos connaissances, votre contexte.
- **Boîte à outils de questions** : obtenir de meilleures réponses grâce à une collection de "formules magiques" (prompts) prêtes à l'emploi.
- **Orchestrateur** : utiliser un framework comme Langchain, qui s'assure que tous ces outils fonctionnent bien ensemble.
- **Gardien éthique** : découvrir les usages qui sont faits de l'IA dans l'entreprise et en garantir un usage éthique.
- **API** : offrir ces services sous forme d'API pour aller au-delà de l'interface de chat et intégrer ces technologies dans votre SI.

IA générative : 4 grands challenges techniques à relever



Conférence scientifique de référence sur l'IA, NeurIPS a dévoilé en décembre 2023 des avancées et des tendances clés pour 2024. **Pour nous, quatre thèmes majeurs façonneront de manière significative l'année à venir.**

Des résolutions de problèmes plus complexes

La technique "Tree of Thoughts" a émergé comme une avancée majeure, permettant une structuration plus fine du processus de pensée. Evolution du modèle actuel de "chain of thoughts", cette nouvelle technique consiste à structurer le processus de pensée du modèle sous forme d'arbre. Chaque branche de cet arbre représente une étape intermédiaire ou une sous-question qui aide à décomposer un problème complexe en éléments plus gérables. Ainsi, au lieu d'aborder un problème dans son ensemble, le modèle suit un chemin logique, examinant différentes possibilités et solutions étape par étape, ce qui permet une analyse plus approfondie et structurée.

Cette approche marque une progression vers des systèmes d'IA plus autonomes, capables de réflexion approfondie.

Des nouvelles architectures de LLM comme Mamba

La plupart des modèles LLM actuels sont fondés sur des Transformers (le T de GPT). Cette architecture a pourtant des limites en termes de longueur de séquence et de coûts d'exploitation. De nouvelles architectures prometteuses sont apparues pour y remédier.

Mamba en particulier révolutionne le traitement des longues séquences de données et offre des performances accrues. Cette architecture ouvre de nouvelles perspectives pour des applications exigeant le traitement de grandes quantités de données. On évoque des séquences pouvant aller jusqu'à 1 million de tokens !

Des IA autonomes et adaptatives

De nouveaux modèles et outils tels que Toolformer et Voyager soulignent la tendance vers des systèmes d'IA plus autonomes. Toolformer apprend à utiliser des outils externes, tandis que Voyager explore et acquiert des compétences dans des environnements virtuels, démontrant une capacité d'adaptation et d'interaction enrichie.

Des méthodologies d'évaluation des capacités des modèles

Les récentes études sur les LLM mettent en lumière la nécessité d'une analyse plus rigoureuse de leurs capacités cognitives. En se concentrant sur la planification et les cartes cognitives, ces recherches révèlent des lacunes notables dans la compréhension conceptuelle des modèles avancés tel que GPT-4. Les auteurs appellent à la création d'une méthodologie d'évaluation réaliste, soulignant que le chemin vers une IA authentiquement intelligente est parsemé d'obstacles méthodologiques et conceptuels.

#Tech #GenAI #Benchmarks

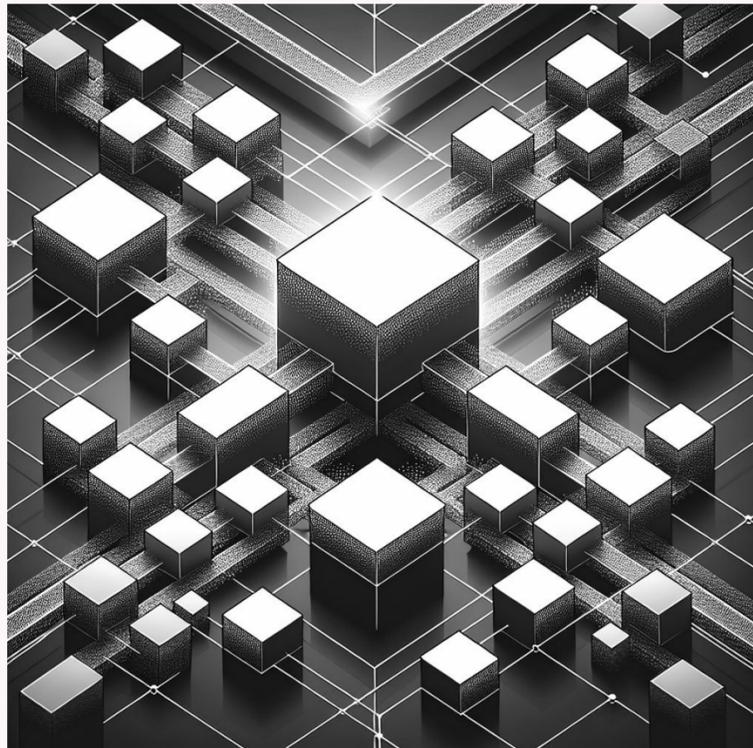
Les benchmarks LLM à suivre en 2024



Les benchmarks de LLM sont des jeux de données librement accessibles au public, conçus pour évaluer les performances et les capacités de ces modèles. Ces benchmarks sont créés par des chercheurs ou par des organisations et sont disponibles pour quiconque souhaite évaluer ses propres LLM. Ils ont 4 finalités :

- **L'évaluation des performances**

Ils fournissent une méthodologie standardisée pour mesurer à quel point les LLM se comportent efficacement dans une variété de tâches, telles que la réponse à des questions, la création de résumés de texte, la traduction, etc.



- **La comparaison des modèles**

Ils permettent aux chercheurs et aux développeurs de comparer différents LLM entre eux, en identifiant leurs points forts et leurs faiblesses.

- **Le suivi de la progression**

Ils servent à suivre l'évolution des performances des LLM et à mettre en évidence les domaines qui nécessitent des améliorations.

- **Assurer l'équité et la transparence**

En mettant à disposition du public ces benchmarks, d'autres utilisateurs peuvent ainsi reproduire et valider les résultats.

Il existe plusieurs manières d'évaluer des LLM : évaluations sans exemple préalable (on parle de "zero-shot learning") ou avec exemples ("few-shot learning"), celles fondées sur l'interaction humaine, etc., ainsi que différents critères selon lesquels leur évaluation est faite (créativité, interactions avec d'autres modèles, compréhension du comportement humain, etc.).

Open LLM Leaderboard offre un large panel intéressant de benchmarks. TriviaQA, HumanEval, GSM8K et MMLU restent autrement des ressources pertinentes et sûres pour aborder la question de l'efficacité des LLM.

#Tech #GenAI #LMM

Du LLM aux LMM - les modèles multimodaux

Le multimodal : l'IA générative pour l'analyse des documents courants

Les LLM sont spécialisés dans la donnée textuelle. Ils rencontrent donc des limites lorsqu'il s'agit d'interpréter des données multimodales, qui comprennent non seulement du texte mais aussi des images, des sons et des vidéos. L'avenir de l'IA générative réside ainsi dans l'utilisation de modèles multimodaux.

Les documents d'entreprise contiennent souvent des éléments visuels comme des images, des graphiques et des schémas. Ces éléments sont riches en informations contextuelles essentielles à la compréhension du document. Les modèles multimodaux, dits LMM, permettent de combler le fossé entre le traitement du langage naturel et la perception visuelle.

En combinant la puissance des LLM avec la capacité de traiter des données visuelles, ces modèles promettent :

- Une extraction d'informations plus efficace : les modèles multimodaux peuvent extraire automatiquement des informations à partir d'images et de textes, ce qui rationalise la gestion des documents et rend les informations plus accessibles.
- Des générations plus riches : en intégrant des données visuelles, les modèles multimodaux peuvent générer des textes plus complets.
- Des recommandations plus personnalisées : dans le domaine du commerce électronique, ces modèles peuvent fournir des recommandations personnalisées en tenant compte des images, des descriptions de produits et des évaluations des utilisateurs.

Exemples d'applications multimodales améliorées par les LMM : rédaction automatique de résumés de documents, recherche d'information multimodale, assistance à la clientèle à partir de photos et de texte, etc.



Les applications GenAI multimodales seront dans les roadmaps 2024 des entreprises

#Tech #GenAI #SLM

LLM ou SLM ? Les modèles de langage se mettent au régime



Travailler avec des Large Language Models implique des contraintes notamment de coûts et d'infrastructure. Les Small Language Models (SLM) se démarquent de leurs homologues plus imposants en offrant une approche plus efficiente, flexible et personnalisable. Pour créer un SLM, on a d'abord besoin d'un LLM pour servir de référence (ou de fondation) et, par un processus de "distillation", on parvient à entraîner un modèle plus compact.

L'objectif poursuivi lors de la création d'un SLM est d'optimiser un rapport bénéfices/coûts : on cherche à maintenir un niveau de performance comparable au LLM de référence en réduisant progressivement le nombre de paramètres nécessaires.

Une fois entraînés, ces SLM sont beaucoup plus flexibles car ils nécessitent moins de puissance de calcul à l'usage et occupent moins de place dans la mémoire des appareils qui l'embarquent. Cette approche ouvre de nouvelles possibilités pour décentraliser l'utilisation des modèles de langage en permettant leur déploiement dans des systèmes du quotidien comme nos ordinateurs et téléphones portables. D'autre part, la réduction du nombre de calculs nécessaires pour effectuer des prédictions fait des SLM des bons candidats pour les applications nécessitant du temps réel ou une forte réactivité comme la retranscription vocale.

Des modèles spécialisés, experts dans un domaine

Il est particulièrement difficile de personnaliser un LLM : cette opération peut être instable, nécessite énormément de données de qualité, très coûteuse également... C'est principalement pour cela que si peu d'entreprise sont capables de proposer de réelles alternatives à ChatGPT. Les SLM permettent de faire sauter cette barrière à l'entrée en diminuant toutes ces complexités (proportionnellement à la réduction du nombre de paramètres). On peut donc plus facilement obtenir des modèles spécialisés qui, en moyenne, sont moins bons qu'un modèle généraliste, mais font preuve d'expertise dans leur domaine (un peu comme nous finalement).

Un bénéfice moins connu de cette spécialisation est la diminution des hallucinations : en réduisant le périmètre de compétence, on facilite le choix d'un corpus d'entraînement très pertinent qui limite aussi l'exposition de données sensibles.

#Tech #GenAI #LAM

Le début des agents autonomes



Rabbit r1, pionnier des "assistants de poche", dopé au LLM Perplexity



Dans la famille des déclinaisons des LLM, voici venir les LAM : Large Action Models. L'idée est d'aller au-delà d'assistants "passifs", qui répondent à des requêtes, pour créer de véritables agents "actifs", opérant des actions.

Il n'est toujours pas question de remplacer l'humain, mais bien de l'augmenter, en lui fournissant un assistant capable de réaliser à sa place une suite de tâches répétitives, sans valeur ajoutée ; c'est en quelque sorte une revisite de la promesse du RDA (Robotic Desktop Automation) rendue possible par l'IA générative.

Le rôle des LAM : orchestrer et automatiser les assistants LLM

Le RDA automatisait une suite de tâches précises (ouvrir telle application, créer un utilisateur, noter son ID, ouvrir une 2e application, y entrer cet ID, etc.) à partir d'un input prédéfini. Grâce aux capacités de traitement du langage naturel, l'objectif des LAM est de permettre à l'utilisateur de définir un objectif final et de laisser l'agent déterminer l'enchaînement des actions à accomplir pour y parvenir.

Le tout récent Rabbit r1 en donne un premier exemple (perfectible) dans la sphère personnelle : son ambition est de remplacer le besoin d'interagir manuellement avec les apps d'un smartphone par des commandes vocales en langage naturel (par exemple, réserve-moi un VTC pour aller à telle heure à tel endroit).

Dans le monde professionnel, le LAM viendra automatiser et orchestrer des actions impliquant plusieurs assistants LLM. La priorité est donc à l'industrialisation des assistants, mais il faut avoir en tête cette évolution !

Vectoriser les données de l'entreprise pour dialoguer avec le SI

Agacé par un résultat non pertinent de votre IA générative préférée, vous avez déjà tenté de changer des éléments de votre prompt ? Si oui, alors comme M. Jourdain qui pratiquait la prose naturellement, vous avez eu recours au prompt engineering.

Le résultat fut-il meilleur ? Probablement, mais certainement toujours loin de vos attentes initiales. Vous avez ainsi touché du doigt les limites de "l'ingénierie de la requête". Le prompt engineering est une vraie compétence, utile, mais il serait illusoire de penser 1) les utilisateurs maîtriseront cet art, et 2) que cela résoudra les soucis de pertinence du contenu.

Le vrai sujet : offrir un dialogue avec le SI en langage naturel

Ce qui donnera véritablement de la valeur à vos applications d'IA générative en entreprise, c'est la capacité offerte aux collaborateurs d'accéder au patrimoine informationnel à travers un dialogue en langage naturel. D'obtenir des réponses pertinentes pour leur domaine d'activité, pour leur métier. Le niveau d'optimisation de la requête sera un plus, qui pourra s'affiner au fil du temps.

Cela nécessite de mettre en place du RAG (Retrieval Augmented Generation) : une architecture qui va combiner la puissance des LLM pour la compréhension des questions et la formulation des réponses et le contenu des documents des utilisateurs (RAG sur des PDF individuels) voire des bases d'informations des entreprises.

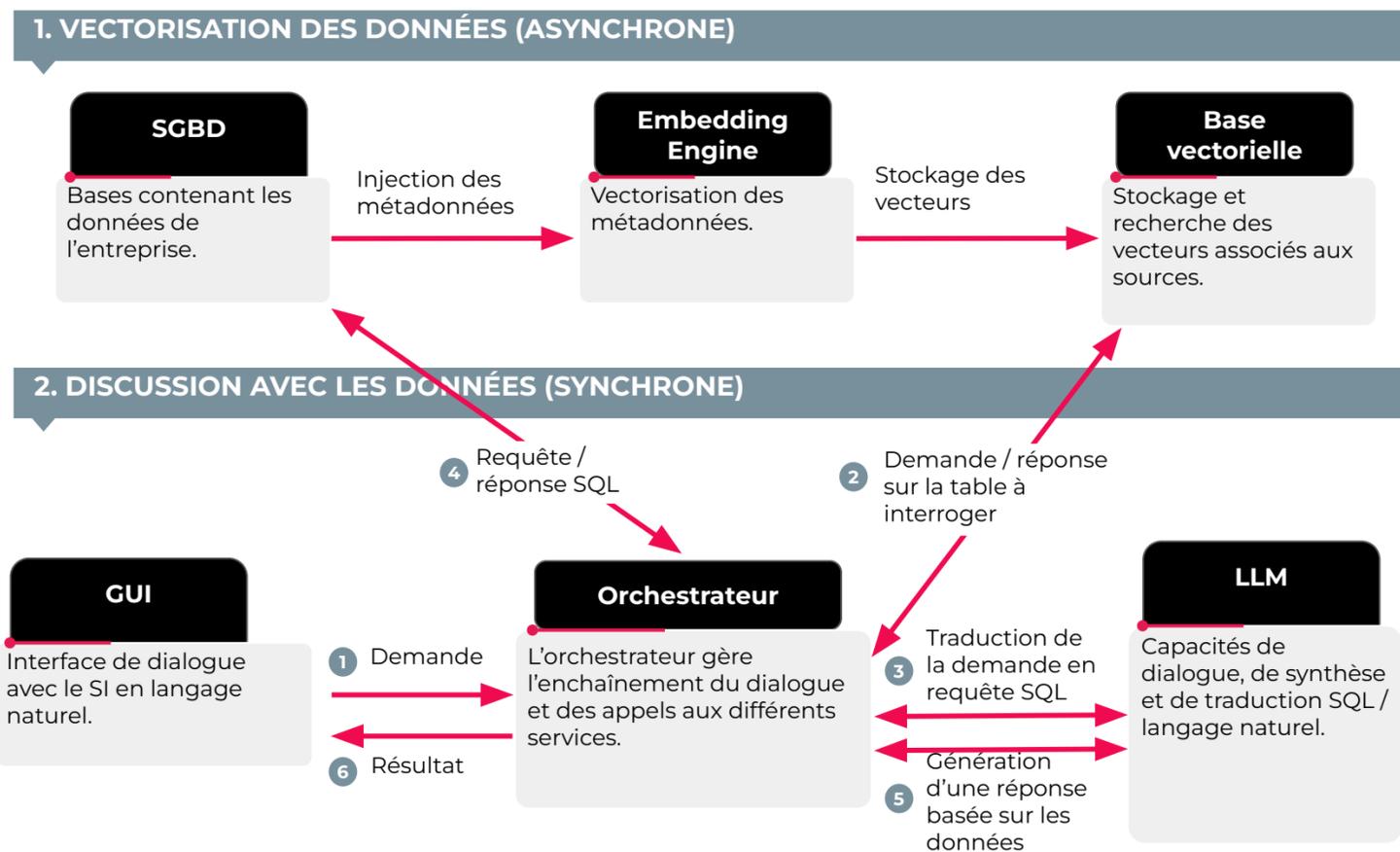
Donner ainsi accès aux données et documents demande de les vectoriser au préalable. Typiquement, vous disposez de dizaines de documents techniques, des PDF de plusieurs dizaines de pages, et vous souhaitez permettre à vos collaborateurs d'interroger cette base de connaissance en langage naturel comme s'ils discutaient avec un documentaliste. Il faut alors passer par une phase dite d'embedding : la création de vecteurs représentant leur contenu, que les moteurs d'IA générative sont capables d'interpréter.

Embedding de documents et données structurées

Cette vectorisation du contenu de l'entreprise est aussi réalisable pour les données structurées : chiffres de vente, nombre d'employés, indicateurs RSE, etc. C'est là où la promesse de l'IA générative en entreprise devient exceptionnelle. C'est là également que le prompt engineering prend tout son sens : selon l'habileté qu'on aura à formuler des prompts pour dialoguer avec le SI, les résultats pourront être très différents.

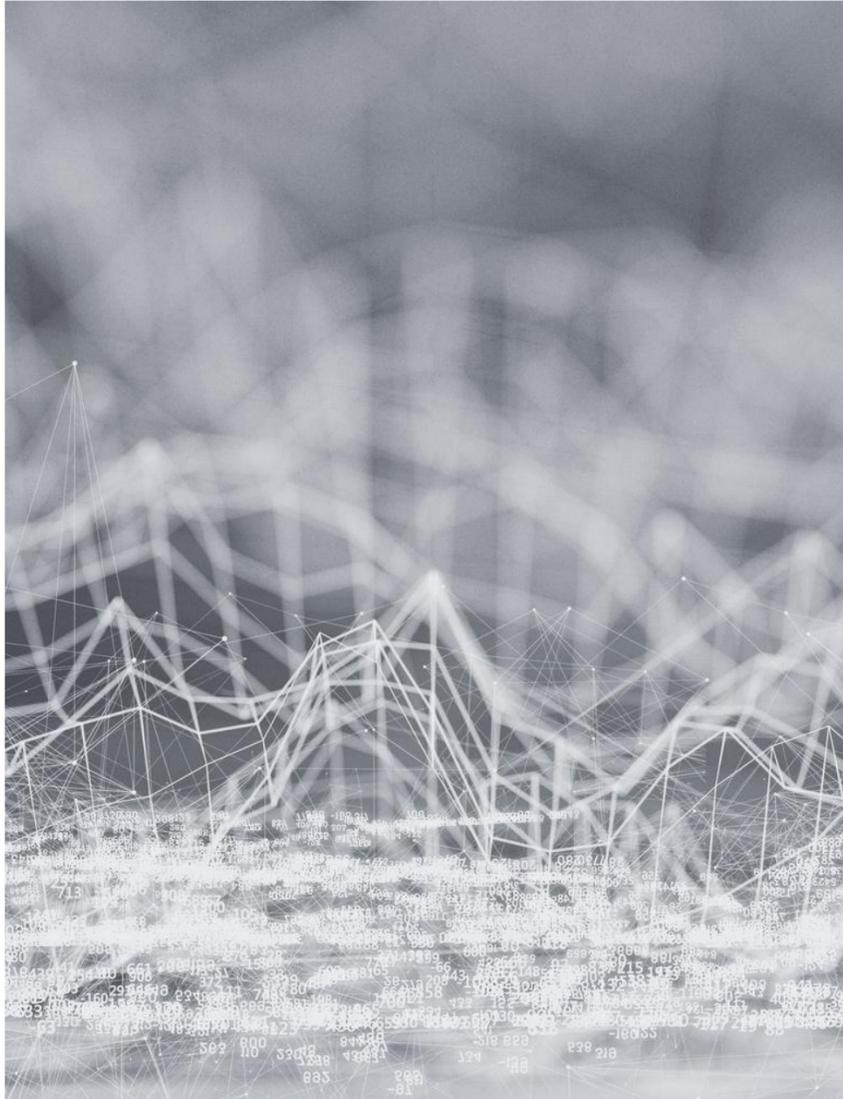
D'où l'importance de coupler ces nouvelles interfaces de chat avec des bibliothèques de prompt : des questions peaufinées, dont on sait qu'elles généreront des résultats pertinents.

Bien entendu, cela demande de prendre quelques précautions. Les bases de données doivent être correctement renseignées et les métadonnées intelligibles. Il faut trouver le bon découpage des documents (en "chunks" analysables par les LLM). Il faut aussi se prémunir contre des risques de sécurité (SQL injection). Mais très clairement, cette approche de requêtage en langage naturel est idéale pour de très nombreux utilisateurs qui pourront s'appuyer sur un assistant personnel pour naviguer au sein du patrimoine informationnel de l'entreprise.



#Data #DataConcepts #GenAI #Prompting

L'art du Prompt



Le rôle du prompt, ou de l'instruction donnée à un modèle d'IA générative, est devenu un élément-clé de nos interactions avec ces systèmes. Que ce soit à travers les interfaces utilisateur ou par l'écriture directe de prompts par les développeurs, le prompt est la pierre angulaire qui permet de tirer parti de la puissance des LLM pour diverses tâches.

“Prompt Engineering” versus “Prompt ingéniosité”

Le prompt se divise en deux catégories distinctes : les prompts générés par les utilisateurs des interfaces et les prompts écrits par les développeurs. Dans le premier cas, les utilisateurs interagissent avec des modèles pré-entraînés en fournissant des instructions en langage naturel. Dans le second cas, les développeurs interagissent directement avec le code en écrivant des prompts dans un langage informatique. Ce sont eux qui vont itérer sur les prompts afin de trouver les meilleures formulations, qui pourront être mises à disposition des utilisateurs au sein d'une bibliothèque de prompts.

Cela ne veut pas dire que les utilisateurs passant par une interface de chat ne puissent pas faire preuve d'une certaine ingéniosité dans leurs prompts.

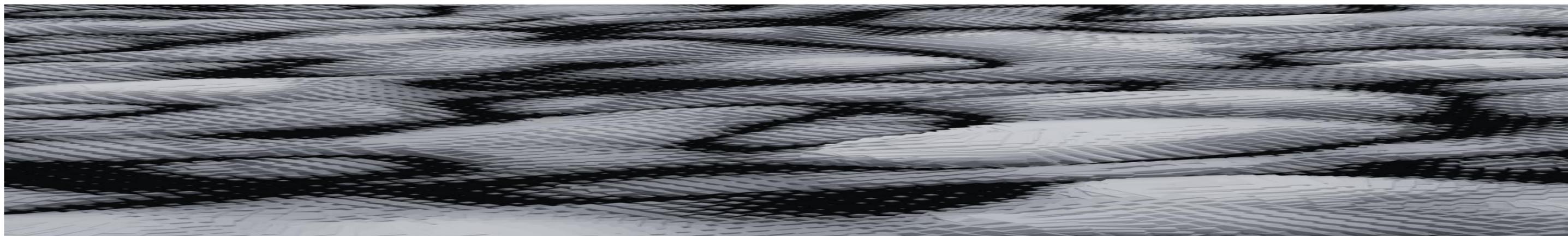
Quelques bonnes pratiques

Le vrai pouvoir du prompt réside dans sa polyvalence. En quelques phrases bien formulées, il est possible d'obtenir un script complet, un résumé d'un document complexe, ou même la génération automatique d'un rapport détaillé. Bien que son utilisation puisse sembler simple, il existe plusieurs bonnes pratiques à suivre pour obtenir des résultats pertinents et efficaces :

- Rédiger des informations claires et structurées
- Donner le plus de détails possibles sur le contexte
- Indiquer le format de sortie désiré
- Expliquer la finalité de la requête

#Data #IA #AgentsAutonomes

L'automatisation à l'ère des LLM : les agents autonomes



Dans le domaine en constante évolution de l'intelligence artificielle, les modèles LLM tels que GPT-4 sont en train de transformer radicalement la notion d'automatisation. Leur impact s'étend bien au-delà des tâches simples, touchant également des processus complexes. **Les modèles LLM donnent une nouvelle dynamique aux agents autonomes en leur octroyant la capacité de segmenter une vaste gamme de problèmes en sous-problèmes plus gérables.** Cette compétence essentielle leur permet de planifier méthodiquement et de construire une suite d'action menant à une solution.

Deux études notables illustrent la progression dans ce domaine. La première, Toolformer, montre que les modèles de langage peuvent apprendre de manière autonome à utiliser des outils externes pour améliorer leurs performances dans une variété de tâches. La seconde, Voyager, présente un agent autonome dans Minecraft, capable d'apprendre et d'évoluer de manière continue sans intervention humaine, en utilisant GPT-4 pour générer et améliorer son propre code et ses compétences.

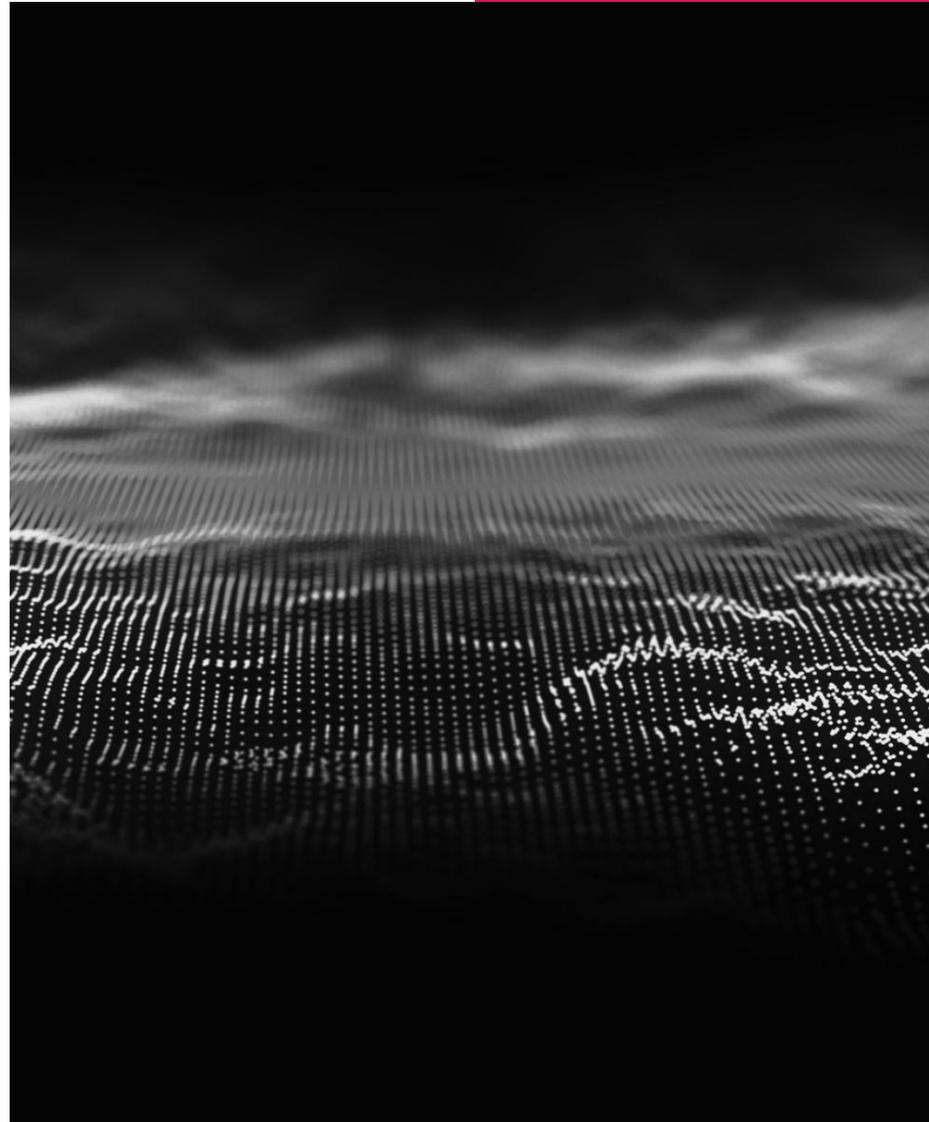
Toutefois, l'intégration de ces technologies avancées soulève des défis significatifs, notamment en matière de sécurité des données et de fiabilité des systèmes. La gestion des informations sensibles par ces agents autonomes exige des mesures de sécurité robustes pour prévenir les abus et les violations de données.

En entreprise, une approche itérative et de l'accompagnement

Dans le contexte professionnel, l'application de ces technologies LLM transforme la manière dont les entreprises opèrent. Ces outils simplifient le traitement de grands volumes de données, facilitant ainsi l'acquisition de connaissances pour les décisions stratégiques. De plus, la capacité de communication en langage naturel des agents autonomes ouvre de nouvelles possibilités pour l'amélioration de l'expérience client et l'efficacité des services. A noter toutefois que cette automatisation accrue amènera la nécessité de former les employés pour qu'ils puissent tirer pleinement parti de ces nouvelles technologies.

La complémentarité des agents et des LLM inaugure une ère d'automatisation et d'innovation. Comprendre et intégrer ces technologies de manière stratégique est crucial pour maximiser leur potentiel au sein des entreprises, tout en abordant de manière responsable les implications éthiques et pratiques. Cette transition vers une automatisation plus intelligente nécessite une approche itérative et bien informée, pour assurer une adoption bénéfique et durable de ces avancées technologiques.

Gouvernance



/06

#Data #DataGouv #Stratégie

Les stratégies data et GenAI à l'épreuve de la valeur

Développer une stratégie data de nos jours est essentiel pour exploiter la valeur des données "dormantes" au sein d'une entreprise, d'autant plus que la démocratisation de la GenAI transforme les cas d'utilisation et l'accès aux données.

Pas de valorisation sans plateforme data

Il est crucial de souligner que l'utilisation des algorithmes GenAI, en dehors de l'expérimentation, requiert une plateforme data. À l'instar de la construction d'une maison, les cas d'utilisation pour valoriser la donnée doivent reposer sur des fondations solides, représentées par une plateforme data.

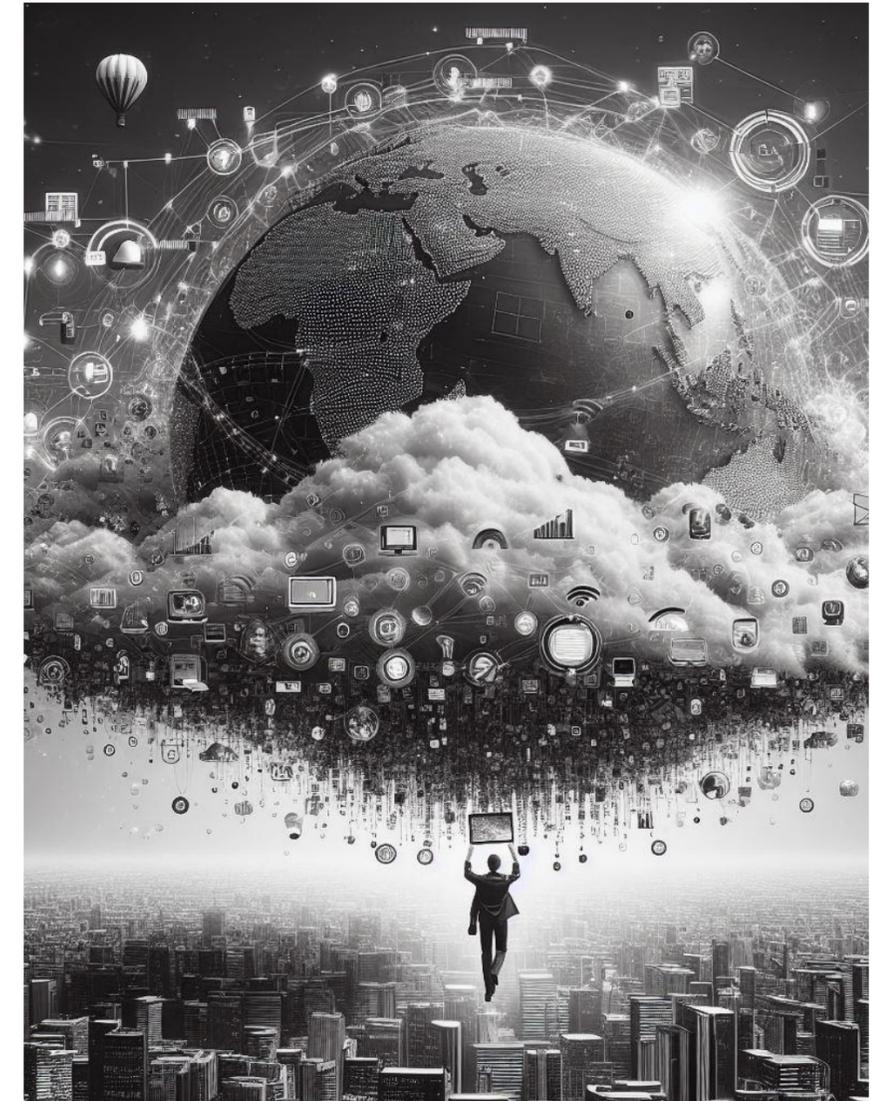
La valeur d'une plateforme : la clé de la pérennité

Dans cette ère de changements technologiques et de conditions économiques incertaines, l'évaluation claire de la valeur d'une plateforme data est primordiale. Les entreprises doivent éviter les écueils classiques liés aux nouvelles technologies :

- ne pas se focaliser uniquement sur l'aspect technique de la GenAI ;
- ne pas en faire un miroir aux alouettes, avec des promesses de millions économisés ou produits en plus ;
- ne pas oublier que la qualité de la donnée en entrée sera déterminante pour le résultat final.

Il est impératif de revenir à la pertinence des cas d'utilisation et de piloter la stratégie data en se concentrant sur la création de valeur. Cette approche revêt une importance particulière, car les algorithmes GenAI, bien qu'ouvrant de nouvelles perspectives, sont gourmands en ressources.

Les concepteurs de plateformes data doivent rationaliser l'utilisation des technologies et définir des KPI clairs quant au coût et à la valeur créée par la plateforme. Ils doivent également intégrer toute la chaîne amont (préparation de la donnée, maintien des bases de connaissance...) dans leurs calculs. Dans une industrie aussi dynamique que la data, seul un pilotage axé sur la valeur permet d'assurer la stabilité et la durabilité d'une plateforme.



#Data #DataGouv #GenAI

L'IA Générative pour améliorer la qualité de la donnée

Actuellement, les algorithmes génératifs sont en passe de révolutionner la philosophie des plateformes de données. Les applications dans le domaine de la data et de l'IA traditionnelle sont nombreuses.

Un exemple significatif est la transformation de la Business Intelligence (BI) grâce à la capacité de ces algorithmes à comprendre les questions liées aux activités business et à extraire les données pertinentes pour y répondre, redéfinissant ainsi les perspectives de la self-BI. La capacité d'interagir avec des données structurées et non structurées, via des algorithmes multimodaux ou la combinaison de modèles, permet une vision simplifiée des données du SI. Toutefois, il ne faut pas se leurrer : le passage à l'échelle de tels cas d'usage est complexe, en grande partie du fait du travail que cela demande sur la préparation de la donnée.



La génération de la donnée au cœur d'une plateforme plus fiable

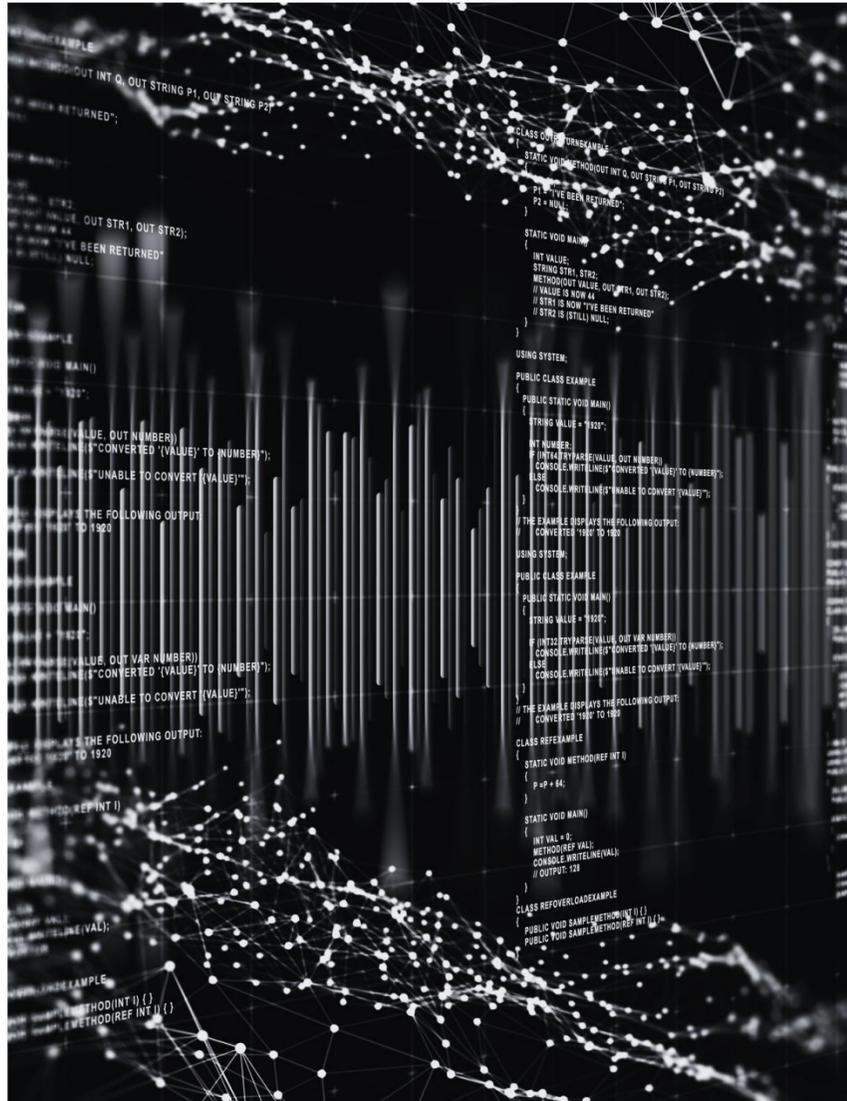
Il faut garder en tête cependant que l'IA générative peut elle-même contribuer à la qualité de la donnée. Les algorithmes génératifs sont capables de détecter des données manquantes mais aussi de créer de nouvelles données à partir de données existantes, ce qui ouvre de nombreuses possibilités dans le domaine de la data. Les données synthétiques peuvent ainsi être utilisées pour générer des jeux de tests pour fiabiliser les pipelines de données afin de s'assurer qu'ils fonctionnent correctement avant d'être déployés en production.

La création de données offre des opportunités prometteuses pour améliorer la gestion des données personnelles. Les développeurs peuvent travailler avec des données générées pour préserver la confidentialité des informations. De plus, la recherche et le tagging automatiques des données permettent un suivi plus efficace et renforcent la sécurité des plateformes de données. En ce qui concerne le développement, la capacité de générer des tests automatiques et une documentation contribue une fois de plus à renforcer la fiabilité des développements.

La donnée générée peut aussi servir de base d'entraînement pour des algorithmes de classification. De nombreux cas d'usage ont émergé autour de la génération d'images, pour entraîner des IA à distinguer des multitudes de formes, états ou variations d'un objet ou d'une situation.

#Data #DataGouv #GenAI #Rôles #ML

Quels rôles pour inclure l'IA générative dans sa stratégie Data ?



L'essor de l'IA générative et des LLM renforce le besoin de déterminer précisément les rôles et responsabilités en matière de data et d'IA, ne serait-ce que pour anticiper les besoins en recrutement et formation. Quelques tendances notables sont à surveiller comme l'augmentation de la demande en expertise en IA générative et en LLM, l'émergence de profils de plus en plus hybrides gardant un lien fort avec les équipes métiers mais bénéficiant de nouvelles capacités augmentées par l'IA, le renforcement de l'importance de l'éthique et de la transparence. Les professionnels de la data et de l'IA seront tenus de redoubler d'efforts pour garantir l'honnêteté, l'impartialité et la traçabilité de leurs modèles.

Voici un aperçu des principales fonctions et attributions associées à chacun d'entre eux :

- Les **Data Analysts** manipulent la donnée pour en tirer des enseignements clés, afin de résoudre des problèmes du business ou d'aider à prendre des décisions mieux informées. Ce rôle est amené à s'éloigner de la DSI pour être plus proche des métiers car les LLM contribuent à faire sauter la barrière technique.
- Les **Data Scientists** construisent des modèles mathématiques de machine learning pour répondre à des problématiques métiers. Dans la majorité des cas, ils s'appuient sur des modèles existants qu'ils personnalisent pour répondre à des enjeux opérationnels, il aussi seront en charge de la veille technologique en l'absence de profils dédiés à la recherche.
- Les **Machine Learning Engineers** se concentrent sur le déploiement et l'opération des modèles d'apprentissage automatique dans des environnements de production. Leur principal objectif est de traduire les modèles développés en solutions logicielles robustes et évolutives. Il doivent être moteur pour la mise en place de la démarche MLOps et bientôt de LLMOps afin de détecter automatiquement toute dérive.
- Les **Data Engineers** développent et maintiennent les outils et infrastructures permettant l'analyse de la donnée. Ce sont eux qui vont implémenter les pipelines de données et permettent sa circulation efficace au sein d'une entreprise.
- Les **AI Research Scientist**, qu'on retrouve chez les entreprises les plus innovantes et ambitieuses, se consacrent à la recherche fondamentale en IA, explorant de nouvelles idées, concepts et approches pour résoudre des problèmes complexes. Leur travail doit faire rayonner l'excellence technique de l'entreprise tant via des prises de parole dans des conférences prestigieuses que par des publications de haut niveau.

Remerciements

à l'ensemble des contributeurs de ces TechTrends, merci à vous de nourrir ces réflexions, au quotidien,

à tous nos lecteurs, merci pour vos feedbacks et vos encouragements,

à tous les clients de SFEIR et WENVISION, qui nous font confiance pour guider leur stratégie et mettre en œuvre nos préconisations !



#TheEnd

TENDANCES DATA & IA

2024

Un grand merci de nous avoir lu.

Vous êtes d'accord ? Vous aimeriez discuter ou approfondir un point ? Contactez-nous, venez en parler avec nous, nous serons très heureux de poursuivre l'échange !

Et nouveauté cette année, nous mettrons régulièrement à jour ce document, afin qu'il reste un document de référence tout au long de l'année.

contact@sfair.com

contact@wenvision.com

[sfair]
WENVISION