



DATA TRENDS

2023

onepoint.
beyond the obvious

DATA TRENDS

2023

édito

Cette seconde édition de nos Data Issues, Data Trends s'inscrit dans la continuité de ce qui fut le déclencheur de l'étude originelle : « *les tendances que nous observons en matière de data et d'intelligence artificielle apportent-elles des réponses concrètes, opérationnelles et opérables aux problématiques rencontrées par nos clients ?* ». Bien sûr l'étude se garde de fournir une réponse tranchée qui n'aurait que peu de sens en dehors de contextes très spécifiques. Cependant, avec désormais une année de recul, quelques conclusions s'imposent.

D'une part les problématiques rencontrées par les clients sont stables dans le temps. Déjà, la première avait mis en lumière une homogénéité des constats en la matière, les organisations de toute tailles et de tous secteurs (y.c. public) semblant rencontrer des obstacles identiques jusque dans leur formulation ; avec un panel étendu (82 Chief Data Officers interrogés), et une année de maturité supplémentaire, les points de douleur demeurent relativement constants.

D'autre part l'effervescence en matière de réponses, notamment technologiques mais aussi méthodologiques et conceptuelles n'a pas ralenti. Plus précisément, de nouvelles tendances apparaissent tandis que d'autres fusionnent ou sont « absorbées » dans les opérations de nos clients. À l'arrivée, nous avons réduit le nombre de tendances, leur donnant davantage de corps, pour n'en retenir que 8, contre 10 en 2022.

Enfin, et c'est la poursuite d'un phénomène observé l'an dernier, les filières data opèrent un retour aux sources, et plus précisément aux fondamentaux les amenant à aller chercher une valeur immédiate. En 2021, les Chief Data Officers nous déclaraient vouloir relever le défi de la gestion de la qualité de la donnée pour retirer davantage de valeur des cas d'usage mis en œuvre ; cette année, le marché semble s'orienter vers des cas d'usage moins ambitieux, moins consommateurs de ressources humaines, financières et technologiques, mais dont la valeur perçue est plus immédiate – bien que moins forte. Cette « fin de l'abondance » est caractéristique de la maturité des filières data.

Ainsi les tendances présentées se constituent en cinq grandes familles : (i) les tendances reconduites à l'identique, (ii) les tendances qui se poursuivent cette année sous forme d'une hybridation avec les tendances identifiées l'an dernier, (iii) les tendances qui disparaissent car elles ont été intégrées dans les opérations et ne sont plus porteuses de nouveauté, (iv) les tendances qui n'ont pas trouvé leur voie et sont donc ajournées, et enfin (v) les tendances qui sont nouvelles, issues de la créativité des expertises et des technologies data.



Gontran Peubez

Partner Data & Intelligence Artificielle

sommaire

01 Méthodologie

06 Méthodologie

02 Data issues

09 Introduction
10 Évolution des problématiques data
17 Nouvelles problématiques data

03 Data trends

21 Évolution des tendances
24 Les tendances 2023
39 Data mesh & data démocratisation
41 Cas d'usage data & IA

04 Ouvertures

44 Confiance numérique
46 ChatGPT et avancées des modèles génératifs

Métho
dologie

Méthodologie

Fidèle à la volonté de recueillir sur le terrain les perceptions des acteurs qui font la data, cette deuxième édition de l'étude « Data Issues, Data Trends » s'appuie sur une série d'interviews (approche qualitative), réalisées entre septembre 2022 et janvier 2023 auprès d'un panel de 82 Chief Data Officers (CDO), issus de divers secteurs d'activités en France.

Contrairement à l'étude précédente, les intitulés de poste des opérationnels interviewés relèvent spécifiquement de l'appellation « Chief Data Officer » : nous n'avons pas souhaité élargir ce panel aux fonctions « non data », qu'elles fussent informatiques ou métiers.

Dans le cadre de ces échanges, nous avons abordé leur situation actuelle sur la data (stratégie, organisation, problématiques, objectifs, ...) ainsi que leur positionnement par rapport à des sujets tendanciels tels que la protection des données, la souveraineté numérique.

Aux bornes de l'exercice de terrain et afin d'en compléter la vision, nous avons, pour cette nouvelle édition, recueilli les perceptions d'une dizaine d'acteurs technologiques ; le nombre auraient pu en être beaucoup plus élevé, mais nous tenions à donner la parole à des acteurs dont nous pouvons constater les réalisations concrètes auprès de nos clients.

L'ensemble de ces informations a permis de mesurer :

- l'évolution des problématiques récurrentes par rapport à 2022 ;
- l'utilisation de la data en matière de business case ;
- l'orientation des entreprises par rapport à des évolutions contextuelles (souveraineté numérique, ...).

À l'issue de cette séquence rythmée par les interviews, les experts data et intelligence artificielle de onepoint se sont réunis pour mettre en commun leur vision : la centaine d'heures d'entretiens réalisés a permis d'esquisser la synthèse des problématiques et des tendances identifiées en 2022 : certaines se sont avérées, d'autres sont suspendues, d'autres encore sont apparues. Cette étude en dresse le portrait pour l'année 2023.

Panel de l'étude

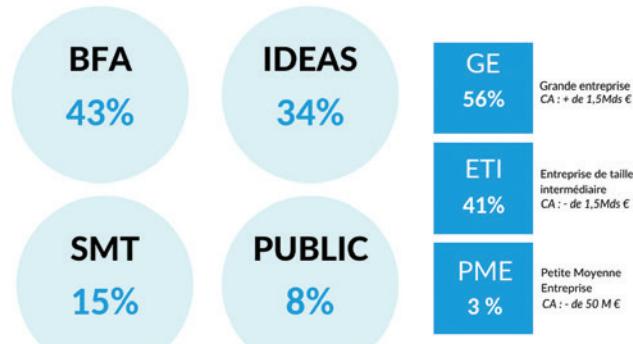
Nos invités

Nous avons eu l'opportunité de compter sur un panel de **82 chief data officers**, de tout horizon.

Indicateur égalité H-F

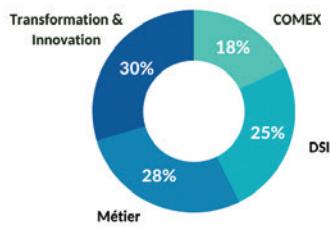
Parmi les 82 CDOs, nous avons 31% de femmes sur cette édition contre 26% l'année dernière, soit un chiffre en augmentation qui témoigne de l'appétit des organisations à nommer une femme à ce poste dans la lignée de "#womenindata".

Domaines et tailles entreprises



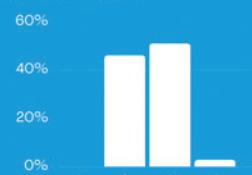
BFA : Banque, finance et Assurance | SMT : Service, Média, Télécommunication
IDEAS : Industrie, distribution, Energie, Automobile, Santé

Positionnement des data office dans les entreprises



Près d'un tiers des data office sont présents dans des directions sur les **aspects transformationnels** (digital, innovation, ...). Ils sont suivis de près par ceux qui agissent au **métier** (28%) où nous constatons plus d'engagement en termes de **business case data et IA**. 25% évoluent dans les DSI et cela offre visiblement plus de **maturité sur l'industrialisation** des initiatives IA. Le reste concerne les DO avec un **rattachement comex** (18%) qui offre plus d'autonomie et de transversalité.

La maturité data



Nous affichons la maturité que les entreprises se sont attribuée entre une note de 1 (peu mature) à 5 (très mature) sur la donnée en général. 51% se sont positionnées sur une maturité à 3 soit des fondations consolidées mais encore des travaux à mener pour être data driven.

Data issues

Data issues

1. Introduction

Si la plupart des organisations ont désormais pris conscience que les données constituent un actif stratégique, leur exploitation demeure pour elles un défi, et en particulier pour leur CDO. Entre aspiration à plus de fluidité pour leurs clients et contraintes opérationnelles, les responsables de filière data doivent trouver leur chemin à l'image de l'un de nos invités de cette édition, « nous aimerais tous que l'accessibilité et la consommation des données soient aussi transparentes et fiables que l'eau qui s'achemine depuis les nappes phréatiques (ex : producteur), en passant par les centres de traitement et canaux de distribution (ex : plateformes data) jusqu'à nos robinets (ex : consommateur). »

En 2021, les interviews réalisées auprès des 58 CDOs nous avaient permis de mettre en évidence 3 grandes familles de problématiques que sont :

- **People** : Dans cette rubrique, nous ne parlons pas de technologie mais de l'être humain qui l'utilise. Ce sont les collaborateurs de l'entreprise, du top management à l'opérationnel, qui rencontrent des points de douleur entre l'appropriation de la culture, les compétences fonctionnelles/techniques disparates et le recrutement difficile de talents sur un marché inflationniste en matière de salaire.

• **Gestion des données** : Cette thématique rassemble les difficultés à maîtriser une donnée qui est par nature changeante, diffuse, ou tour à tour, extrêmement volumineuse ou encore volatile et dont on cherche en permanence à renforcer la connaissance, la qualité ou l'harmonisation et de minimiser les efforts de réconciliation nécessaire. Une maîtrise de bout en bout de la donnée reste le graal pour atteindre la «source unique de vérité»

• **Industrialisation** : La troisième problématique concerne les freins à l'exploitation des cas d'usage data/IA à l'échelle de l'entreprise, encore très peu nombreux pour des raisons technologiques (plateforme, ...), financières (coûts de gestion de la donnée, ...) ou pour encore d'autres raisons, que nous détaillons dans la fiche correspondante.

En 2022, l'ensemble de ces difficultés perdurent à différents niveaux dans les entreprises selon leur maturité générale sur la donnée. Loin d'être fatalistes, la plupart des CDOs interrogés nous partage le « pragmatisme » dont ils font preuve pour aborder sur le moyen ou long terme ces problématiques, tout en s'évertuant à produire de la valeur rapidement auprès de leur COMEX afin de préserver les investissements de leur feuille de route.

Si les trois problématiques identifiées en 2022 perdurent en 2023, nous identifions toutefois une nouvelle famille liée à ce que nous nommons « Fin de l'abondance ». Les programmes data – visant d'ailleurs à résoudre les trois thèmes évoqués ci-dessus – se sont parfois dotés de moyens colossaux. Dans un contexte de frugalité (l'inflation, la guerre en Ukraine et le climat) se généralisant pour l'ensemble des

fonctions de l'entreprise, les CDOs sont invités eux aussi à se concentrer sur l'essentiel, à privilégier davantage les gains rapides que les projets de long-terme au retour sur investissement hasardeux et donc à se concentrer sur les fondamentaux data.

Nous avons donc introduit une quatrième problématique, nommée « Data Fasting », à la faveur d'un jeu de mot mêlant la notion de vitesse et de jeûne et donc de diète.

2. Évolution des problématiques data 2022

Issue #1 – People

Rappel de la problématique

Le sujet PEOPLE traite de l'enjeu de développement des compétences, et surtout de leur mise en action effective au service de la performance de l'organisation, tout autant que de la culture data.

En 2022, la dimension PEOPLE se déclinait selon deux problématiques :

- l'appropriation à tous les niveaux de l'entreprise des enjeux data et l'acceptation des responsabilités correspondantes par les acteurs,
- le recrutement, la fidélisation des profils data et l'animation de leurs trajectoires de montée en compétence.

En 2023, les difficultés de recrutement et de fidélisation persistent pour 100 % des organisations interrogées :

- Les profils sont toujours rares et les exigences salariales grimpent de manière disproportionnée en regard des compétences affichées (« les prétentions des Data Engineers explosent », « il est impossible de trouver des ML Engineers », « c'est une catastrophe sur le régaliens car il n'y a pas de filière de formation dédiée : Data Owner, Data Steward... »).
- Le taux de rotation au sein des équipes est fort, même chez les prestataires externes.
- La compétition pour les profils se joue aussi au sein même des organisations. En effet, les directions métier se dotent de profils data avec en priorité : les Data Analysts, les Data Scientists aux côtés des Data Stewards, Data Owners. Cela renforce le besoin d'une culture data partagée et d'un modèle opérationnel fluide.

Par ailleurs, une définition plus complète des attentes en matière de compétences émerge chez plus d'un tiers des CDOs consultés : un « bon profil data » est un profil hybride. Au-delà de détenir des expertises techniques, il s'agit aussi d'avoir de l'appétence métier, être fédérateur, bon communicant, savoir piloter rigoureusement un projet de bout-en-bout.

En 2023, les incertitudes économiques et politiques pourraient toutefois diminuer la tension sur le marché du recrutement. Cependant, l'attractivité des organisations et leur capacité à proposer des missions enthousiasmantes avec du « sens » resteront au premier plan.

Actions lancées en 2022 pour résoudre la problématique

Chez 95 % des organisations interrogées, des plans de formation data sont proposés, avec un effort de prise en compte de la diversité des profils et leur maturité. La palette des activités s'est remarquablement enrichie d'accompagnements opérationnels, événementiel comme les Café Data ou les Data Days, et d'activités ludifiées (livret d'apprentissage de type cahier de vacances pour COMEX, par exemple). Toutefois, les organisations ont du mal à gérer la diversité et la complexité des besoins, ainsi qu'à intégrer l'amélioration des soft skills dans les parcours.

Sur le plan du recrutement, tout le monde participe, ce n'est plus la chasse gardée des RH et des spécialistes du recrutement. Toutes les organisations interrogées ouvrent et diversifient leurs approches : recrutement sur les réseaux sociaux, par la participation à des salons, ou encore des meet-ups avec un intérêt fort pour trouver l'inspiration auprès des pairs.

Pour pallier la rareté des profils sur le marché, quelques organisations choisissent de recruter des « têtes bien faites » hors des filières data qui seront formées en interne. Cela suppose une projection conjointe sur du moyen-long terme et une vraie capacité d'accompagnement (catalogue de formations, plan de formation personnalisé, mentoring).

Enfin, à la fois pour répondre aux besoins forts de delivery mais aussi pour former sur le terrain via des projets concrets et rendre autonomes leurs équipes avec le soutien d'experts data (externes), les organisations développent des centres d'excellence avec leurs prestataires. La montée en

compétences est notamment reconnue par le biais de l'obtention de certifications qui contribuent à éléver le niveau standard et valorisent les profils.

Où en sont les entreprises de la place ?

L'ensemble des organisations interrogées ont un programme de développement de la culture et des compétences data. Toutes reconnaissent qu'il s'agit d'un chantier complexe et qu'il est de longue haleine. La principale difficulté est de savoir combiner les différents leviers pour être efficient (formation, accompagnement opérationnel, etc.) car, lorsqu'ils existent, ils sont historiquement gérés en silos.

L'acculturation des COMEX se développe et lorsque la prise de conscience est faite, l'organisation gagne en efficience à travers un meilleur alignement et engagement plus fort sur les sujets stratégiques. Toutefois, c'est une démarche qui nécessite qu'au moins un sponsor se mobilise pour la lancer et maintenir l'engagement des décideurs dans le temps.

Enfin, l'enjeu phare aujourd'hui est le déploiement d'un modèle de gouvernance complet avec une responsabilisation des acteurs métier (data ownership), une collaboration transverse et de bout-en-bout dans une logique de Data Product.

Nos convictions pour accélérer

Développer la culture data reste une démarche de transformation profonde de l'organisation. Aussi les initiatives doivent avant tout poser le bon diagnostic pour ensuite accélérer sur des bases solides.

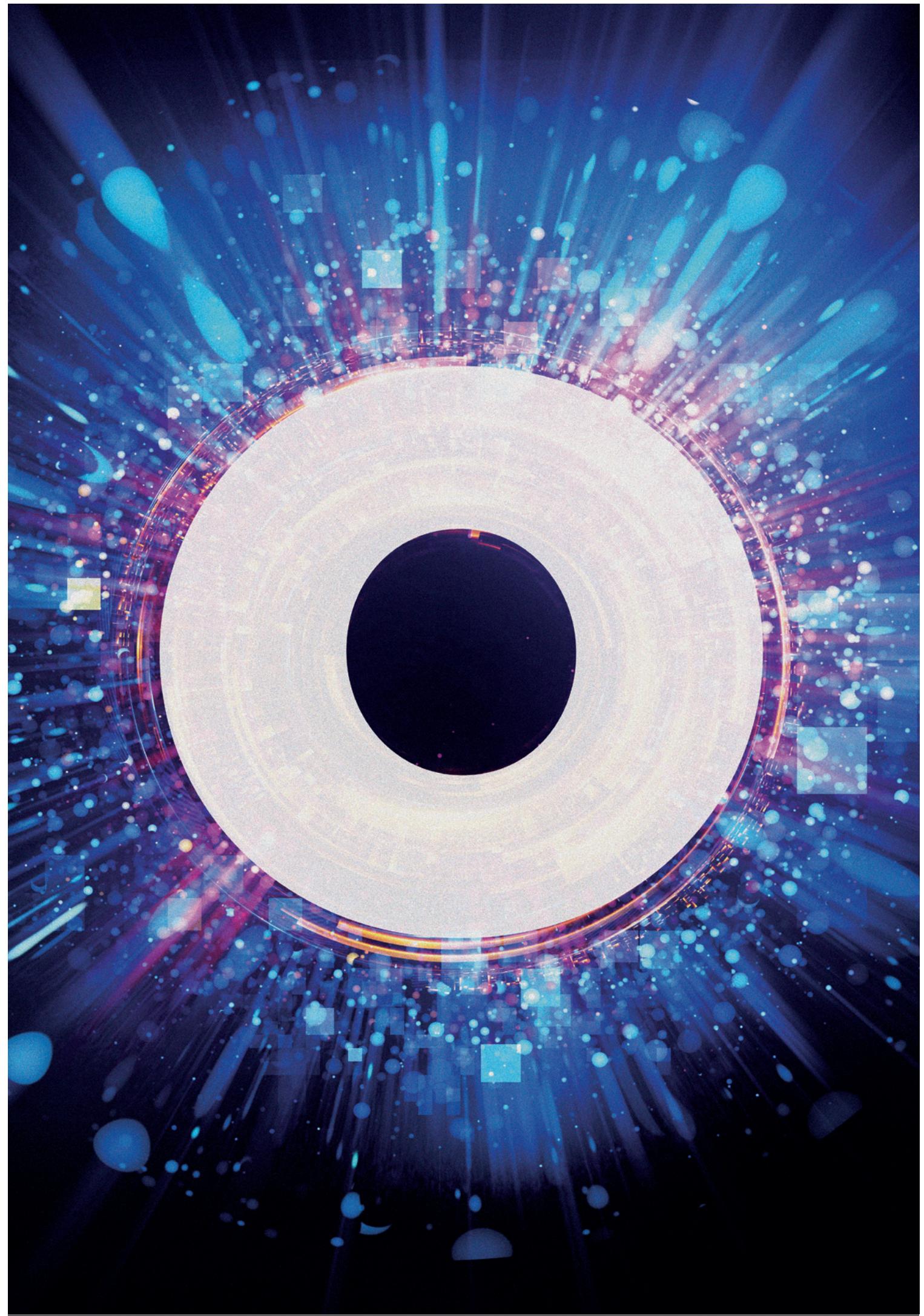
- Faire un véritable état des lieux des valeurs, pratiques, normes et comportements à tous les niveaux hiérarchiques. Investir dans des moyens d'accompagnement à la montée en compétence : formations, mentoring, coaching, accompagnements opérationnels sur les projets.
- Utiliser les initiatives métiers stratégiques comme déclencheur pour développer la culture et les bonnes pratiques data : les efforts d'apprentissage sont mieux justifiés et acceptés lorsqu'il y a un enjeu d'entreprise clair (engagement RSE, plateformisation).
- Développer l'expérience collaborateur au-delà de la data (modernisation des SI, qualité des équipes, projets et outillages data), favoriser la transversalité, et être transparent (ne pas survendre le poste, la maturité de l'organisation, l'état du SI, les moyens...).

Issue #2 – Gestion des données

Rappel de la problématique

Les problèmes et défis de la gestion des données ont évolué au fil du temps et sont restés très importants malgré le développement rapide de nombreuses solutions technologiques et méthodologiques, et ce en raison de l'augmentation constante de la quantité et de la complexité des données collectées et exploitées.

En 2022, la dimension gestion des données soulevait plusieurs problématiques : le manque de connaissance et maîtrise des données qui circulent dans l'entreprise, une



perception défaillante de la qualité des données exposées et une carence dans la gestion des données transverses (données de référence).

En 2023, la qualité de données demeure un enjeu majeur pour 80 % des CDOs impliqués dans l'étude.

- La qualité des données est souvent pilotée en bout de chaîne dans les couches de reporting, avec des remédiations ponctuelles.
- Malgré la prise de conscience de la valeur de la donnée, le partage d'informations entre les métiers n'est pas fluide voire inexistant.
- L'explosion des données semi structurées et non structurées, le besoin croissant de collecter les données externes (ex : ESG) et la complexité des systèmes d'information rendent difficile le croisement et l'analyse des données.
- Il est toujours aussi difficile d'avoir des personnes responsables pour maintenir dans le temps une donnée de qualité.

Le caractère continu de cette activité, l'évolution constante et la diversité des problématiques de gestion des données restent des défis majeurs qui nécessitent que les responsables mettent en place des stratégies et des processus pour les gérer de manière efficace.

Actions lancées en 2022 pour résoudre la problématique

Pour 95 % des CDOs interrogés, la priorité en 2022 était de poser les fondations de la gouvernance de données ou stabiliser son déploiement pour mettre à disposition des

données de qualité et en adéquation avec les usages. De nombreuses approches ont été mises en œuvre pour accélérer ces chantiers parmi lesquelles :

- Recruter un responsable de la gouvernance des données pour sortir de la théorie et ancrer la gouvernance by design.
- Démarrer par les domaines de données transverses, considérées comme des données « essentielles » pour l'entreprise (ex : données clients) afin de déployer la gouvernance et de démontrer la valeur ajoutée du partage de la donnée.
- Outiller la démarche en rendant les outils faciles à utiliser (ex : data catalog), responsabiliser les métiers et l'IT sur la documentation de l'ensemble des données et la qualité de données attendue.
- Co-construire avec les métiers des pipelines de données qui mettent à disposition des datasets certifiés avec des données gouvernées de bout en bout et documentées à un bon niveau de granularité pour l'entreprise.
- Créer et animer une communauté data s'engageant dans la durée autour d'un but commun.

L'enjeu 2023 est de maintenir les travaux en cours qui visent à ancrer opérationnellement la gouvernance et passer ainsi à l'échelle.

Où en sont les entreprises de la place ?

Les entreprises sondées ont pratiquement toutes lancé le chantier gouvernance de données que ce soit en cadrage, planification ou exécution, et démontrent un avancement

significatif sur certaines familles de données notamment les données de risques et financières avec des résultats tangibles.

Le principal défi des entreprises aujourd’hui est de généraliser la gouvernance pour exploiter tout le potentiel de leurs données. En effet, les difficultés restent de démontrer la valeur ajoutée de la gouvernance des données qui est encore perçue comme un centre de coûts, de rendre les outils faciles à utiliser pour favoriser leur adoption auprès des métiers et de les embarquer (engagement fort) sur la durée.

Néanmoins, tous les secteurs d’activité n’ont pas la même maturité et ne sont pas au même niveau de déploiement : Il s’agit des secteurs réglementés (à l’instar du secteur bancaire) qui sont plus avancés et sont désormais à une étape d’urbanisation des données (usages, contrôles qualité, accès à la donnée partagée et unique).

Il est indéniable que le chemin reste encore long pour beaucoup, malgré les efforts consentis pour y parvenir.

Nos convictions pour accélérer

La data gouvernance doit être augmentée par l’IA et intégrée by design (cf. Trend 3 – Beyond data governance), ce qui implique de :

- Promouvoir une culture de partage de données dans l’entreprise et créer des processus de partage de données standard.
- Être pragmatique en tirant la gouvernance par les usages et l’adaptant par domaine de données ce qui permet de le faire au fil de l’eau tout en créant de la valeur pour les consommateurs.

- Adopter et s’adapter à de nouveaux patterns comme le Data Mesh qui imposent la gouvernance de données pour être efficient.
- Inclure du point de vue de l’outil les briques nécessaires à la bonne interopérabilité avec le reste du système d’information.

Issue #3 – Industrialisation

Rappel de la problématique

L’industrialisation est le processus visant à déployer en production des solutions Data et IA. Il nécessite souvent une infrastructure technologique avancée, des compétences spécialisées et une culture d’entreprise axée sur les données afin de garantir l’apport de valeur des cas d’usage en permettant leur passage à l’échelle.

Si l’industrialisation des chaînes de traitement BI est aujourd’hui globalement maîtrisée par les entreprises, les principaux défis auxquels elles sont confrontées concernent les projets de Data Science et d’IA.

Après une vague d’expérimentations, les premiers cas d’usage d’IA ont commencé à être mis en production ces dernières années. Néanmoins, leur proportion reste faible et en 2022, les entreprises continuent de rencontrer des freins à l’industrialisation.

Parmi ces freins, se retrouvent majoritairement les décalages entre les équipes issus des DataLab et celle de l’IT qui accentuent la frontière entre les processus et les technologies d’expérimentation d’une part et de production d’autre part. Les équipes DataLab sont généralement composées de Data Scientists et d’experts

en IA, qui sont chargés de l'exploration des données et de l'expérimentation de nouvelles solutions dans des environnements « bac à sable » qui s'affranchissent des normes et standards de la production ainsi que de ses contraintes de performances et de sécurité. Les équipes IT, quant à elles, sont responsables de la mise en place des solutions opérationnelles, de leur maintenance et de leur évolution. Cette division peut conduire à une incompréhension entre les deux équipes et ralentir l'industrialisation des projets d'IA.

Actions lancées en 2022 pour résoudre la problématique

En 2022, les organisations ont travaillé à mieux structurer l'industrialisation des projets en adressant en priorité les enjeux technologiques et de compétences :

- Déploiement de démarches MLOPS à travers la mise en œuvre de plateformes data science de bout en bout ou d'architectures sur-mesure. Celles-ci permettent de gérer l'ensemble du cycle de vie des projets d'IA, de la collecte des données à l'analyse des résultats, en passant par la conception des modèles et leur déploiement.
- Recrutement de profils ML Engineer en capacité d'être responsables de la conception des modèles de Machine Learning, de la collecte et du traitement des données, de la mise en œuvre de l'infrastructure de déploiement, ainsi que de la maintenance et de l'amélioration de la solution.

Une acculturation progressive du Management aux enjeux de l'industrialisation a également permis de soutenir ces

initiatives en acceptant des délais de mises en marché plus longs sur les premiers cas d'usage au bénéfice d'une accélération future une fois les modalités de mise en œuvre éprouvées.

Où en sont les entreprises de la place ?

En 2023, la tendance à l'industrialisation des solutions de données et d'IA se poursuit et se renforce, avec une prise de conscience croissante de la nécessité de définir des modèles opérationnels complets pour accompagner cette transformation. Les entreprises comprennent désormais que l'industrialisation ne se limite pas à des questions techniques, mais nécessite également des réflexions approfondies sur l'organisation, la gouvernance et les processus.

Les modèles opérationnels sont des cadres de travail qui permettent de structurer les différentes étapes de l'industrialisation des solutions de données et d'IA, en définissant des rôles et des responsabilités clairs pour chaque membre de l'équipe, ainsi que des processus et des normes à respecter. Ces modèles opérationnels permettent de projeter l'industrialisation dans une vision pérenne et favorisant l'amplification de son empreinte, avec toujours plus de produits data industriels sur davantage de périmètres.

Nos convictions pour accélérer

Nous identifions deux leviers pour accélérer face à l'enjeu d'industrialisation.

Tout d'abord, mettre en place des dispositifs de type « AI Factory » adaptés au développement et à la maintenance de solutions IA industrielles :

- Ces structures organisationnelles permettent de rassembler des experts en Data Science, des ingénieurs IT, des spécialistes en métier et en produit dans un même environnement de travail. Cette organisation favorise la collaboration entre les différents métiers, la compréhension mutuelle des enjeux et la mise en commun des compétences, ce qui permet d'accélérer le développement de solutions de données et d'IA industrialisées.
- Les « AI Factories » s'appuient également sur des environnements techniques normés et sécurisés pour garantir la qualité et la fiabilité des solutions développées. Ces environnements techniques sont conçus pour répondre aux besoins de l'entreprise en matière de collecte, de stockage, de traitement et d'analyse des données, ainsi que pour assurer la sécurité des données.

Ces dispositifs sont conçus pour permettre aux cas d'usage d'apporter de la valeur dès la phase de POC (Proof of Concept), en garantissant une approche « industrielle by design ». Cela signifie que les développements techniques sont conçus dès le départ pour être industrialisés, en utilisant des normes et des standards reconnus, afin de garantir une maintenance facilitée, une évolutivité maîtrisée et une scalabilité optimisée.

Mais aussi, prendre en compte l'organisation du run des solutions d'IA dès la phase de conception et de développement :

- La mise en production d'une solution d'IA nécessite une organisation spécifique pour garantir sa qualité et sa fiabilité dans le temps. Cette organisation doit être anticipée en amont pour éviter des

dysfonctionnements, des retards ou des coûts supplémentaires.

- Il est recommandé de définir dès le début les processus de run et de monitoring des modèles en production, en impliquant les équipes IT, les équipes de Data Science et les responsables métier. Les équipes doivent travailler ensemble pour identifier les métriques de performance pertinentes, mettre en place les processus de contrôle qualité et de validation, ainsi que pour élaborer des plans d'urgence en cas de dysfonctionnement.
- L'utilisation des outils de gestion de configuration et de gestion de version permet de suivre l'évolution des modèles et garantir la cohérence de la solution en production.

3. Nouvelle problématique data identifiée pour 2023

Issues #4 – « Data Fasting »

Constat / diagnostic

En 2023, la plupart des organisations ont démarré leur voyage dans la data depuis plusieurs années (une majorité depuis plus de 5 ans, un petit tiers depuis plus de 2 ans). La rentabilisation des premiers investissements n'est au mieux que très partiellement démontrée et il est paradoxal de constater que malgré un long chemin restant à parcourir certaines entreprises tendent à faire du sujet data une composante banale de la production

informatique – avec les aléas budgétaires associés.

Par ailleurs, l'attention s'est tournée vers des cas d'usages d'intelligence artificielle qui relèvent des projets attractifs, innovants, séduisants. Cette belle vitrine masque derrière un voile d'innovation la lourdeur des travaux de data management et de remédiation, les collectifs qu'il faut savoir mobiliser, les enjeux de maintenance des modèles algorithmiques.

Pourquoi ça ne marche pas ?

La transformation data n'est pas finie, au contraire ! Combien d'organisations peuvent dire sans mentir que la data est un de leur levier de performance aujourd'hui ?

À l'instar des DSI qui ont opéré leur mue vers une position de Business Partner, les organisations data doivent piloter un budget de run et un budget d'investissement avec une focalisation forte sur le ROI. Ce changement de mindset de Créeateur à Bâtisseur se produit alors que les fondamentaux data (rôles, process, outils) sont encore mal appropriés dans bien des cas. Quelques signes des temps : des transformations SI mal articulées avec les plans stratégiques business (« les métiers ne comprennent pas bien les bénéfices de la data platform »), le déploiement des rôles et responsabilités dans les organisations sans envisager d'ouverture de poste (« d'accord pour le data ownership mais on n'a pas de bande passante »), des efforts en data management mal soutenus (« notre data catalog, c'est Excel et Powerpoint »).

Il y a, en effet, une complexité pour les organisations à manier le court terme et le moyen-long terme lorsque la stratégie

data et son plan de route ne sont pas rigoureusement arrimés aux objectifs stratégiques business. Si la valeur de la data est explicite au travers d'un data product, elle naît dès les phases amont de cartographie et de documentation. Cette valeur de connaissance et de maîtrise du patrimoine, souvent cachée, s'avère difficilement mesurable. De même, les efforts de maintien de la qualité des données et des produits data sont considérés comme une charge récurrente sans comptabilisation de la valeur générée par l'intensification de l'usage des données rendue possible ou par la réutilisation des modèles algorithmiques déployés, ou encore plus simplement la transformation culturelle qui l'accompagne.

Pris entre des travaux de long terme sur les fondamentaux data et l'exigence de ROI à brève échéance, les CDO n'ont ainsi pas d'autre choix que de repenser l'organisation des moyens et la production des data products pour aller vers plus de rapidité et d'efficience, tout en gardant le cap sur le développement de la culture data et la poursuite du déploiement du modèle de gouvernance.

Les nouveaux défis pour les CDOs sont donc la vitesse (délivrer de la valeur quasi immédiate) et la frugalité (délivrer de la valeur sans investissements colossaux) : c'est l'objet du Data Fasting qui est la synthèse de ces efforts conjugués.

Nos convictions pour résoudre la problématique

Maîtriser frugalité et vitesse passe avant tout pour le CDOs par la capacité à manier le court et le long terme afin de trouver l'alignement d'objectif parfois

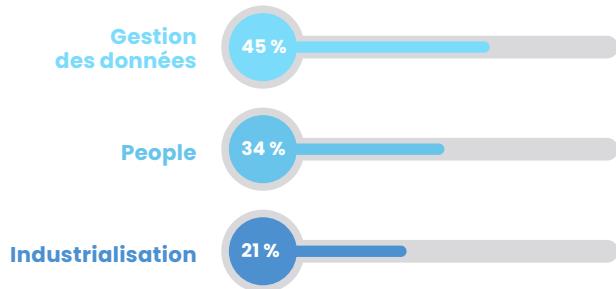
difficilement compatibles entre eux. Pour bâtir les fondations Data sans se lancer dans des projets « cathédrales », pour servir des victoires rapides riches en bénéfices immédiats pour les métiers, nos recommandations sont les suivantes :

- Continuer à renforcer l'engagement et la sagacité des décideurs pour aller au bout des transformations data, avec les niveaux d'investissement adaptés autour du triptyque Business (viabilité)/Technologie (faisabilité)/People (adoption).
- Aller vers plus de simplicité et d'évidence de la valeur : en combinaison avec les initiatives stratégiques profondes et transformantes, disposer d'un volant de projets identifiés avec les métiers pour soutenir l'adoption des bonnes pratiques par l'ensemble des parties prenantes tout en démontrant la valeur des services délivrés par les données (fiabilité des études, amélioration de la capacité d'analyse et de connaissance client/marché, prédiction ou prescription en soutien des commerciaux et conseillers de vente, etc.). C'est le sens des démarches de Data Productization (cf. Trend 1)
- Persister à travailler la data dans toutes ses formes et incarnations pour favoriser l'appropriation et le ROI : mesurer et mettre en qualité des datasets fréquemment utilisés, rationaliser et optimiser la BI pour aller jusqu'au data story telling, traiter des « petits cas d'usages » et valoriser tout cela au sein de l'organisation. « Faire avec », c'est assumer un certain état d'hétérogénéité des solutions technologiques mises à disposition, reconnaître les résistances au changement qui n'excluent pas des ambitions de valeur autour de la data.

Finalement, c'est adopter un mindset de frugalité (cf. Trend 7 - Data Ecosystem Frugality) qui invite à la prise de recul pour mettre en place plus de modularité et d'agilité, et prendre son indépendance vis-à-vis des dogmes qui ont guidé les premiers pas des organisations data.

Interrogés au sujet de leurs priorités de résolution des problématiques 2022, les CDOs se positionnent comme suit :

La première place est trustée par la gestion des données pour 45 % des CDOs qui en font leur chantier prioritaire et clé de voute pour débloquer et exploiter la valeur de leur patrimoine informationnel. La dimension People occupe la deuxième position avec près de 34 % en continuité des travaux déjà engagés pour enrayer le manque de culture et de compétences. L'industrialisation arrive en dernière position avec 21 %.





Data
trends

Data trends

1. Évolution des tendances 2022

Après avoir interrogé les CDOs, et en tenant compte de nos observations du terrain, le panorama de l'évolution des tendances est le suivant (**schéma ci-contre**).

Autour de ces cinq familles identifiées, nous vous proposons dans cette partie de dresser une rétrospective pour comprendre l'évolution par rapport à notre publication précédente.

1.1 Tendance ajournée

Une seule tendance est concernée par cette « mise en pause ».

Data Office as Profit Center

Souvent mise en avant par les CDOs interrogés, force est de constater que sa concrétisation demeure incertaine. Les entreprises semblent encore chercher la recette pour faire du Data Office un centre de profit centré sur la création de valeur par la co-construction avec les métiers. Les difficultés opérationnelles concentrent le rôle du Data Office sur des activités plus opérationnelles pour l'heure. Pour autant, transformer les équipes data office en

centre de profit sera un véritable enjeu de développement au sein des organisations pour les années à venir.

Notre conviction est que cette tendance n'a pas pu se réaliser par manque de maturité, c'est pourquoi nous la considérons comme ajournée jusqu'à une prochaine réactivation.

1.2 Tendances intégrées dans les opérations

Deux tendances : la « Data Academy » et la Data XOps, ont été adoptées dans les opérations quotidiennes.

Data Academy

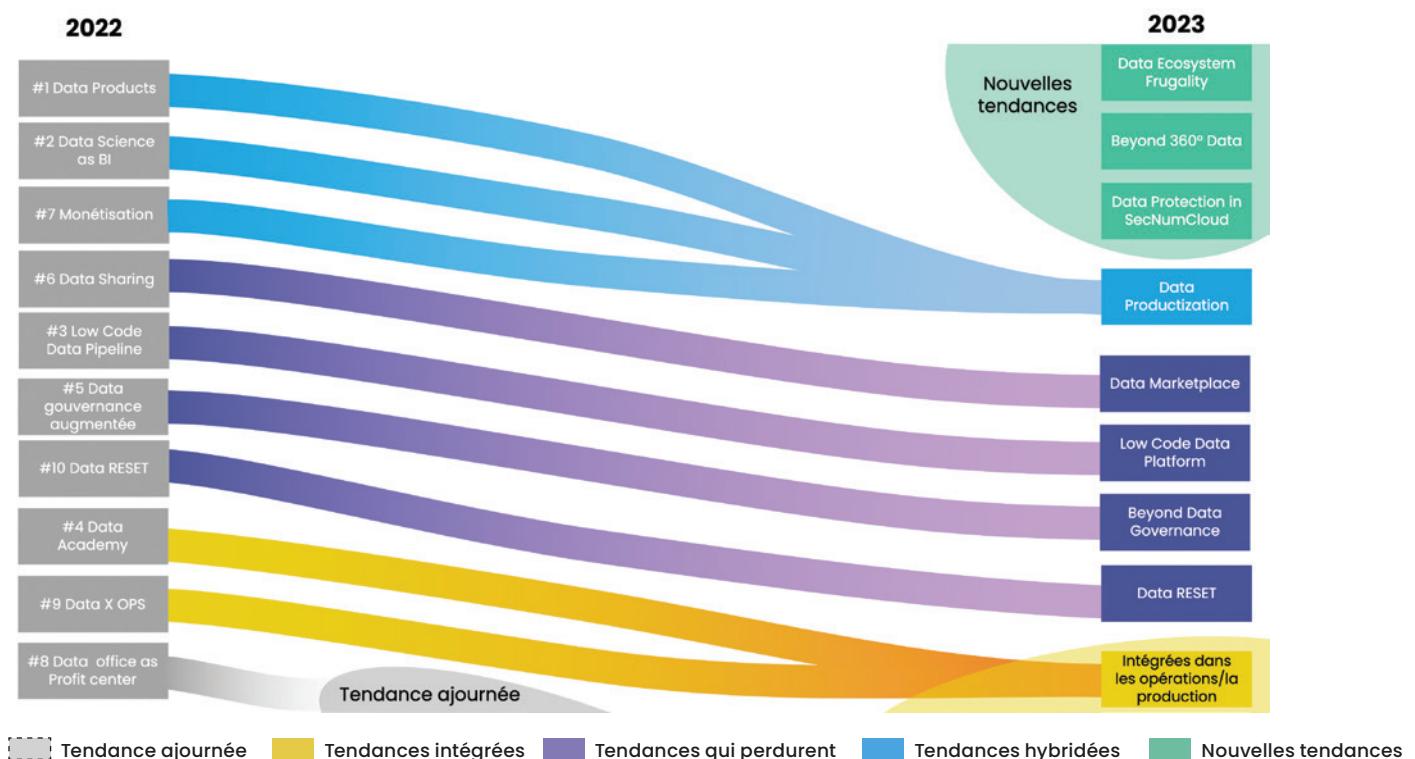
Ce n'est plus une nouveauté, tout le monde en parle, tout le monde en fait. Elle fait désormais partie du quotidien opérationnel des entreprises à travers de nombreuses initiatives mises en œuvre pour acculturer et former les collaborateurs sur la data

et les solutions implémentées. C'est une démarche qui s'inscrit dans le temps pour une transformation profonde, pour instaurer une « Data Academy » complète et passer à l'échelle.

Data XOps

Fréquemment observée et bien ancrée en 2022, la démarche DevOps appliquée aux données s'avère moins complexe qu'il n'y paraît si l'entreprise a déjà développé la culture DevOps. En revanche, lorsque les approches Data x Ops ne sont pas suffisamment orientées vers la gestion de pipelines de bout en bout, les innovations et adaptations technologiques demeurent importantes pour exploiter pleinement le potentiel de cette tendance.

Considérant le nombre de réalisations dans les entreprises et le niveau d'utilisation des outils et principales méthodes, nous estimons que ces deux tendances font désormais partie du quotidien.



1.3 Tendances amplifiées ou transformées

Quatre tendances se sont retrouvées en 2022 au centre de préoccupations nouvelles et ont été particulièrement débattues tout au long de l'année pour prendre une nouvelle direction.

Data Sharing

Une tendance de fond qui permet de casser les silos data, une tendance qui touche à la fois à la culture, l'organisationnel et l'outil. Elle est bien suivie, reste d'actualité tout en évoluant pour donner naissance à une tendance plus englobante qui va au-delà de la problématique d'accessibilité des données en adressant à la fois la fédération des assets de l'entreprise et le self-service data pour créer de la valeur et accélérer la transformation (cf. Trend Data Marketplace).

Low Code Data pipeline

Tendance très bien accueillie en 2022 malgré la question de gouvernance des usages qu'elle soulève et le risque du shadow IT qu'elle peut engendrer. Elle n'a pas été concrétisée dans son ensemble car perçue comme un moyen d'autonomiser les métiers sur les activités exploratoires. L'objectif est désormais de construire des pipelines de bout en bout en intégrant la gouvernance by design comme prérequis pour contrôler les usages self-service.

Data Gouvernance Augmentée

En 2022 la priorité des entreprises sondées était de poser les fondations de la gouvernance et/ou stabiliser le déploiement

opérationnel. La data gouvernance augmentée demeure une tendance sur 2023 car les enjeux des entreprises en matière de gouvernance restent le passage à l'échelle et de faire du dictionnaire de donnée une bible d'entreprise.

Data RESET

Data RESET (Responsabilité Environnementale, Sociétale, Économique et Technologique en matière de Données) est au cœur des stratégies et propositions de valeur des entreprises. Cette tendance a été fortement plébiscitée sur 2022, notamment pour tendre vers la sobriété numérique et l'IA éthique. 2022 a été une année de prise de conscience, de sensibilisation et de réflexion sur la tendance qui intègre entre autres green data/frugalité. Cette tendance qui est une priorité pour les entreprises sondées reste dans l'agenda 2023 qui se veut être une année de concrétisation.

Les initiatives constatées, l'accroissement de l'engouement autour de ces thématiques, confirment notre perception et nous encouragent à affirmer que ces tendances, bien que déjà identifiées en 2022, seront davantage mises en avant en 2023.

1.4 Tendances fusionnées

Les tendances suivantes ont évolué de sorte à fusionner en raison de leurs interdépendances fortes.

Data Product

Le data Product est au cœur des stratégies data des entreprises. Cette tendance a été la plus plébiscitée en 2022. Elle a été amorcée

par bon nombre d'entreprises (en cadrage ou en conception) notamment en tant que produit générique pour contrebalancer les silos data, ou pour s'orienter vers une logique Data Mesh. En 2023 cette tendance continuera à se concrétiser dans sa globalité en évoluant pour permettre d'adresser à la fois les besoins analytiques et opérationnels des organisations (cf. Trend Data Productization).

Data Monétisation

Moins évoquée comme une priorité par les CDOs que nous avons rencontrés, cette tendance semble trop éloignée des priorités des décideurs ou du modèle économique de certaines entreprises et demeure regardée avec perplexité. Les potentielles contraintes juridiques et le risque de perte de confiance vis-à-vis des clients finaux semblent atténuer la volonté des entreprises de travailler sur la question à court-terme. Il devient opportun de lever les freins en adoptant une démarche de privacy by design.

Data science as BI

Bien que la nécessité d'allier du prédictif à la BI pour favoriser la prise de décision soit bien comprise et perçue par tous, cette tendance peine à s'ancrer dans les entreprises. Certaines d'entre elles n'ont pas beaucoup de cas d'usage data science et d'autres s'échinent à industrialiser leurs cas d'usage data science. En sus, l'absence de ressources et compétences clés ainsi que les problématiques de qualité de données voire organisationnelles ne permettent pas de passer aisément le cap, ce qui permet d'imaginer l'avenir de cette tendance comme devenant un produit data à part entière.

En raison d'écart d'appréciation motivés par le manque de cas d'usage, de fortes craintes ou par l'inadaptation des concepts sous-jacents, nous estimons que la « Data Monétisation » et la « Data Science as BI » seront absorbées par la tendance Data Product qui s'amplifie et permettra de déboucher vers une monétisation des produits de données en les transformant en actif valorisable. Cette tendance devient Data Productization.

1.5 Les tendances qui font leur apparition en 2023

Trois nouvelles tendances font leur apparition dans cette nouvelle édition et sont détaillées dans la section suivante. Il s'agit de :

- Data Ecosystem Frugality
- Beyond 360° Data
- Data protection in SecNumCloud.

2. Les tendances 2023

Pour cette nouvelle édition, nous avons décliné le même mode opératoire en nous basant sur notre vision et notre expertise du marché pour vous proposer 8 tendances qui vont potentiellement guider les prochaines orientations sur la data en 2023.

Les tendances 2023 en un clin d'œil

1. #DataProductization - Tendance reconduite, évolution de Data Product.
La Data Productization est la capacité d'une organisation à offrir des produits basés sur la donnée (Data Products).



2. #DataLowCodePlatform – Tendance reconduite, évolution de LowCode Data Pipelines

Les plateformes de données Low code font référence à un ensemble d'outils et capacités qui permettent de construire des pipelines de données de bout en bout (ingestion, stockage, transformation, exposition) avec très peu de lignes de code en s'appuyant sur des langages « universels » connus de tous (ex : SQL).

3. #BeyondDataGovernance – Tendance reconduite

Beyond Data Governance permet d'aller au-delà de la gouvernance défensive en rendant le cadre de gouvernance

adaptatif et en augmentant le data management avec l'IA pour une découverte automatique des données techniques et métiers ainsi que leur lignage.

4. #DataMarketplace – Tendance reconduite, évolution de Data Sharing
Plateforme en ligne qui permet aux consommateurs d'accéder de manière industrielle (en quelques clics) à des données de confiance (in fine des produits), dans une logique collaborative et self-service.

5. #DataProtection in SecNumCloud –
Tendance nouvelle

Capacité désormais offerte aux organisations de disposer d'un niveau de protection éprouvée par l'ANSSI tout en bénéficiant de la flexibilité et de la scalabilité du Cloud.

6. #Beyond 360° Data – Tendance nouvelle

Capacité à acquérir les données d'un objet métier en dépassant les frontières et aller au-delà d'une simple vision à 360°.

7. #DataEcosystemFrugality – Tendance nouvelle

Approche de gestion des données basée sur l'ouverture technologique, la modularité et interopérabilité.

8. #DataRESET – Tendance reconduite

Responsabilité Environnementale, Sociétale, Économique et Technologique en matière de données.

Trend #1 – Data Productization

Pourquoi / Problématique

Alors que la majorité des organisations est dorénavant convaincue qu'il est nécessaire de faire levier sur la donnée pour améliorer sa performance économique et renforcer son avantage concurrentiel, l'utilisation de la donnée, autrefois principalement circonscrite à la sphère analytique de l'entreprise, s'invite jusque dans les processus opérationnels, implantée sous la forme de systèmes d'intelligence artificielle. Pour y parvenir, les filières responsables doivent élaborer des moyens de consommer la donnée, sous des formes variées (tableaux de bord, algorithmes, etc...).

L'heure est à présent à l'accélération et au passage à l'échelle. Comment introduire ces modules data à haute valeur ajoutée

dans l'ensemble des processus analytiques et opérationnels de l'entreprise ? Comment parvenir à contourner cette complexité intrinsèque aux projets data, qui en ralentit la diffusion ?

Définition

Les Data Products, ou « produits data », sont des produits dont l'essentiel de la valeur ajoutée est constitué par la donnée. La Data Productization est la capacité offerte par une organisation de consommer les données sous forme de Data Products.

Un Data Product peut être créé dans le domaine métier responsable des données qu'il valorise. Il peut adopter différentes formes (fichier plat, API, base de données), proposer différents niveaux d'engagement (matérialisés par un SLA - Service Level Agreement). Il est mis à disposition de l'ensemble de l'entreprise via une plateforme en self-service qui en assure la promotion, l'information et la souscription.

En quoi cela répond à la problématique ?

Le Data Product est un véhicule particulièrement adapté à la transformation data de l'entreprise et au passage à l'échelle.

L'adopter procure de nombreux bénéfices. Fonctionnellement, il permet d'aborder de multiples besoins (analytique, self-service, opérationnel). Sur le plan organisationnel, il permet de bien délimiter et ancrer les responsabilités de chacun à l'égard des données et de leur valorisation. Sur le plan de la gouvernance, l'approche produit permet de réaliser progressivement le travail nécessaire, en priorisant selon les besoins des utilisateurs. La gouvernance fédérée permet

de garantir la qualité, l'interopérabilité et la pérennité des offres. Economiquement, le design d'un Data Product pensé pour la réutilisation, permet de bénéficier d'un coût incrémental quasi-nul.

Généralisation / Cas d'application

Dès lors il est possible d'imaginer des Data Products modularisés. Des plus simples (un fichier à plat déposé sur un « quai de déchargement ») aux plus complexes (un module qui calcule pour un prospect les probabilités d'achat pour chaque produit de l'entreprise dans les 2 prochaines semaines), cela entraîne naturellement certaines évolutions. Par exemple, la volonté de disposer de produits réutilisables pousse vers la simplicité d'une part et la traçabilité et la responsabilisation des producteurs d'autre part.

Limites

Parmi les écueils les plus fréquemment identifiés par les early adopters, nous trouvons l'adoption de la culture produit. Il est vrai que passer d'une approche projet à une approche produit nécessite de « changer de logiciel ». Il faut penser « marché des utilisateurs internes » plutôt que « demandeur » et penser problème plutôt que solution. Autrement dit, faire de la recherche produit. Il faut concevoir le produit pour une réutilisation par d'autres utilisateurs, c'est à dire définir son périmètre minimum et renoncer à une adéquation parfaite aux besoins du premier utilisateur (overfitting du produit à son premier client). Enfin, il faut gérer un cycle de vie produit bien plus long que celui d'un projet car il faut faire vivre les produits au gré des évolutions des besoins

des clients (et également des non-clients).

Une autre difficulté réside dans le modèle économique qui sous-tend cette organisation, car dès lors qu'un Data Product Manager travaille à la création et au perfectionnement d'un Data Product au bénéfice d'autres domaines métier, se pose la question de l'équilibre économique de chacun des domaines.

Trend #2 – Data Low Code Platform

Pourquoi / Problématique

Les usages data se diversifient et se multiplient de manière exponentielle, mais les entreprises sont confrontées à un manque de ressources techniques, à une hétérogénéité des compétences à l'échelle de l'entreprise, et parfois à une fuite des sachants ce qui crée des difficultés pour maintenir ou faire évoluer les usages.

La réalité montre que les métiers souhaitent toujours plus d'agilité avec une volonté de permettre à chacun de prendre des décisions intelligentes, et de pouvoir répondre plus rapidement aux attentes des différents consommateurs de données. De son côté, l'IT doit répondre à des enjeux spécifiques : une explosion des données et des sources associées avec une difficulté pour les rendre accessibles, lutter contre le Shadow IT, garantir la scalabilité des solutions tout en gardant les coûts sous contrôle.

Définition

Les plateformes de données Low Code font référence à un ensemble d'outils et capacités

qui permettent de construire des pipelines de données de bout en bout (ingestion, stockage, transformation, exposition) avec très peu de lignes de code en s'appuyant sur des langages « universels » connus de tous (ex : SQL). Ces plateformes accélèrent la production de valeur et suscitent un grand enthousiasme dans la sphère data et contribuent à la mise en place d'architectures de type DataHub / Lakehouse pour fluidifier le lien entre producteurs et consommateurs de données, tout en offrant une véritable approche de self-service pour les différents domaines métiers.

En quoi cela répond à la problématique ?

Les plateformes low-code permettent à des populations « non expertes » de construire et de partager de manière simple des produits de données et de déverrouiller l'accès à la donnée.

L'enjeu est de permettre à des domaines métiers de se concentrer sur le design du pipeline et la valeur produite, tout en apportant des bénéfices tangibles tels que :

- Accélération du Time to Market pour répondre à des problématiques métiers, notamment en phase d'incubation / prototypage (nombre réduit de lignes de codes).
- Abstraction de la complexité technologique pour permettre aux utilisateurs de se focaliser sur l'apport de valeur escomptée et d'utiliser facilement les données.
- Flexibilité accrue et modularité apportées aux domaines en fournissant des accélérateurs « homologués » réutilisables,

avec une capacité à s'appuyer sur des puits de données existants (pour ne pas dupliquer la donnée).

- Industrialisation simplifiée avec les équipes IT et une meilleure maintenance dans le temps du fait de s'appuyer sur des langages universels déjà connus par les ressources de l'entreprise.
- Mise en place d'une gouvernance « native » forte permettant de lutter contre le shadow IT, notamment via le processus d'homologation et l'application des best practices de software engineering (séparation dev/prod, unit tests, CI/CD...).

Généralisation / Cas d'application

Les outils Low Code, intégrant des capacités « by design » de lineage, mapping, transformations et d'orchestration, permettent d'adresser différentes propositions de valeur :

- Fournir un outil favorisant la productivité et offrant une expérience optimale pour les utilisateurs, dans le but d'accélérer la création d'innovation et de nouveaux modèles d'affaires.
- Supporter le nouveau paradigme Data Mesh en offrant les capacités self-services nécessaires pour les domaines.
- Réduire la dette technologique et rationnaliser la manière d'implémenter des pipelines data (batch ou streaming) ETL/ELT.
- Appliquer une démarche de gouvernance by design où les templates / accélérateurs sont homologués, tout en encourageant l'inner sourcing pour pousser à plus de réutilisation et capitalisation dans

l'entreprise (« bout de code », template de pipeline à instancier, ...).

Limites

Ces plateformes ne permettant pas d'adresser des problématiques complexes, leur positionnement et leur intégration doivent faire l'objet d'une étude afin de les positionner dans l'écosystème de données :

- Identifier les populations et les langages ciblés via ce type d'outils.
- Éviter les risques de vendor locking en privilégiant l'utilisation d'outil ou framework open source.
- Définir le cadre commun pour encadrer les usages et mettre en place le bon niveau de gouvernance adapté à ce type d'approche.

Trend #3 – Beyond Data Governance

Pourquoi / Problématique

Le déploiement de cadres de data management s'est généralisé avec la mise en place des Data Offices ; en particulier ces derniers ont rendu possible la gestion des cas d'usages réglementaires.

La généralisation de la démarche en vue de valoriser voire monétiser les données reste un défi. En effet, les démarches réalisées sur un périmètre réduit, i.e. pour des cas d'usages réglementaires, n'ont pas permis l'identification des leviers à actionner pour passer à l'échelle.

Les Data Officers doivent désormais augmenter leur data gouvernance de capacités leur permettant de traiter des cas

d'usage opérationnels métiers porteurs de valeur (en particulier dans la logique de Data Productization évoquée en tendance n° 1) et ce dans des univers informationnels mouvants rendant rapidement obsolètes les initiatives de gouvernance « déclaratives ».

Définition

Au-delà d'une gouvernance des données défensive, la data gouvernance augmentée repose sur le déploiement d'un cadre de gouvernance intelligent (outillé et adapté aux usages) mettant en œuvre des moyens traditionnels épaulés par l'IA by design pour la mise en place d'une démarche à chaque nouveau cas d'usage.

Dès lors, la proposition de valeur de la data gouvernance augmentée est de rendre adaptatif le cadre de gouvernance et de permettre un passage à l'échelle en automatisant la découverte des données techniques et métiers et ainsi que leur lignage.

En quoi cela répond à la problématique ?

Étendre la zone de maîtrise de la gouvernance des données permet de répondre à de nouveaux enjeux métiers ou à des enjeux réglementaires au-delà des périmètres d'application actuels. Cet accroissement de périmètre ne peut se faire à dispositif et outillage constant au sein des Data Offices. L'équipement de la Data Gouvernance permet d'outiller les tâches de documentation particulièrement fastidieuses et rapidement dépassées compte tenu de leur trop rares mises à jour.

L'approche by design s'inscrit dans la limitation des efforts de déploiement en embarquant la gouvernance des données dans des activités

de cadrage ou de spécifications. L'ambition ROIste est par nature portée par le projet. Pragmatisme, itération et progressivité sont les maîtres mots de la démarche.

L'utilisation d'IA permet la découverte de données d'intérêts, pour d'autres domaines que le domaine d'origine qui ne les auraient pas identifiées comme telles, tout en proposant une gouvernance, des règles d'utilisations, etc...

Généralisation / Cas d'application

La mise en place d'une gouvernance augmentée est particulièrement adaptée dans une approche fédérée pour laquelle les données sont gouvernées localement par des principes fédéraux. Domaine après domaine et cas d'usage par cas d'usage, la gouvernance augmentée se déploie au-delà de ses prérogatives défensives. Projet, maintenance, réingénierie, etc... toutes les initiatives sont éligibles à la démarche, d'autant que, plus elle est déployée, plus elle est bénéfique. Plus il existe de données gouvernées, plus les moteurs d'IA pourront en identifier de nouvelles.

Limites

La gouvernance des données nécessite une documentation de qualité, liant l'ensemble des couches du SI (de métier à physique). Les réponses technologiques visant à automatiser cette démarche semblent timides à date. Parmi les explications, la maturité des Data Offices eux-mêmes est à prendre en considération : souvent focalisés sur les cas d'usage réglementaires, rares sont ceux qui visent une gouvernance offensive, déployée de manière adaptative sur différents domaines métiers.

Si les organisations ne sont donc pas

toutes prêtes, il est néanmoins possible d'observer un repositionnement des Data Offices s'engageant dans la voie du Domain Ownership. La montée en maturité, au regard de la gouvernance des données, bat une nouvelle fois les cartes de l'organisation et c'est l'une des opportunités porteuses de ROI dont les entreprises doivent s'emparer.

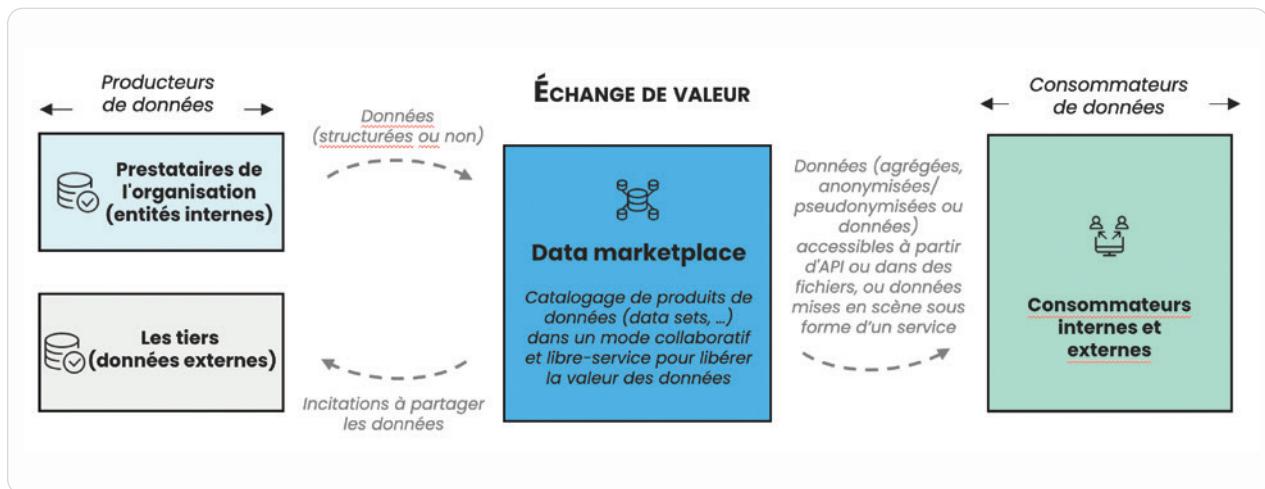
Trend #4 – Data Marketplace

Pourquoi / Problématique

S'il est indéniable que les données sont le carburant des cas d'usages dont la valeur est tirée de l'exploitation de ces dernières, les infrastructures et les organisations actuelles ne permettent pas de mettre en œuvre ces initiatives. La Data Productization (tendance n° 1) rend possible la « fabrication » de produits modulaires dont la valeur provient de la donnée. Le partage et la « commercialisation » de ces produits nécessite une place de marché permettant de les échanger entre producteur et consommateur, selon des modalités convenues (langage commun, interopérabilité, sécurité, niveau de service, ...).

Définition

Une Data Marketplace (place de marché de données) est une plateforme en ligne qui permet aux consommateurs d'accéder de manière industrielle et simple (en quelques clics) à des données de confiance (in fine des produits), dans une logique collaborative et self-service. Ces Data Products sont mis en lignes par des producteurs de données qui s'engagent à respecter le niveau d'exigences de la Data Marketplace. Ainsi la Data Marketplace donne accès en toute confiance à l'utilisateur



métier à des données qui proviennent de domaines au-delà du sien, participant à la démocratisation des données.

En quoi cela répond à la problématique ?

La Data Marketplace est un outil performant au service des directions data pour activer une accessibilité de la donnée à l'échelle de manière industrielle avec cinq caractéristiques principales :

- Un partage de valeur à travers des données de confiance en quelques clics – ou produits (produit d'un producteur ou bien d'un data broker agrégeant différents produits pour en recréer un).
- Un approche self-service, notamment en s'appuyant sur une expérience sans couture pour faciliter la réutilisation et collaboration : accès facilité et unifié à l'information, suggestions personnalisées ; l'ambition est ici de s'affranchir de la complexité d'accès au travers des couches technologiques sous-jacentes.
- Un langage commun, permettant une interopérabilité (non seulement

sémantique mais aussi technique) au sein d'un écosystème de données.

- Un niveau de sécurité des données (respect de la conformité, etc...) et des algorithmes (performance, éthique...).
- Une capacité de suivre la performance des usages des assets (données internes/externes) et de leur ROI.

Généralisation / Cas d'application

L'intégration et la généralisation de l'usage d'une Data Marketplace tirent divers aspects positifs pour l'entreprise bénéficiaire, avec entre autres les possibilités de :

- Créer de nouveaux produits et services issus de plusieurs domaines avec une expérience unifiée et optimisée.
- Distribuer et valoriser voire monétiser la donnée en fonction des consommateurs ciblés (intra-entité, inter-entité ou externe).
- Mettre à disposition les données open data à l'échelle de l'entreprise, en rationalisant les coûts d'acquisition.
- Gouverner les accès by design en simplifiant l'application des politiques

d'accès via la souscription.

- Promouvoir une logique d'écosystèmes de données (par grand domaine métier) pour soutenir de nouveaux enjeux économiques tout en s'orientant vers des modèles à impact positif.

Limites

Malgré les atouts de la Data Marketplace, son succès repose avant tout sur des fondations solides (gouvernance adaptée notamment au niveau sécurité, approche « data privacy » pour l'anonymisation, ...) et une hygiène de tous les instants des producteurs pour assurer la qualité des assets de données mises à disposition via ce canal.

Il convient également de prendre en compte l'environnement technologique de l'entreprise, car la marketplace nécessite une maîtrise et une maturité permettant l'orchestration de la mise en marché, de la productization et de la consommation. En outre, la donnée dans la marketplace est un « corner » parmi d'autres (API, service IT, services métiers, ...), ce qui implique de définir une stratégie globale de marketplace afin de garantir une cohérence d'usage de ces services.

Trend #5 – Data Protection in SecNumCloud

Pourquoi / Problématique ?

Le Cloud (dont les services sont majoritairement offerts par des prestataires extra-européens) et la protection des données ont longtemps semblé impossible à concilier. La protection des données, présente

dans la loi CNIL 1978 a été promulguée 40 ans plus tard au travers du RGPD qui renforce et unifie la protection des données pour les individus au sein de l'Union européenne.

Identifié dès 2008 par les principaux cabinets d'analyse de tendance, le Cloud Computing s'est désormais imposé auprès d'organismes publics ou privés pour héberger et traiter leurs données. Si les acteurs ont d'abord été séduits par la capacité de stockage et de mise à disposition de machines virtuelles (IaaS), les fournisseurs de services Cloud (CSP) ont rapidement proposé des systèmes d'exploitation et des outils d'infrastructure permettant d'offrir de véritables plateformes technologiques (PaaS), avant de proposer des applications logicielles intégrées (SaaS).

Les atouts du Cloud (scalabilité, capacité d'accélération et résilience) n'ont pourtant pas eu raison de la frilosité de nombreux acteurs à héberger leurs données dans le Cloud Public et ces organisations ont utilisé le Cloud au mieux comme un complément d'infrastructure. Aujourd'hui le Cloud offre une alternative crédible aux Data Centers des entreprises, grâce notamment à la certification de sécurité SecNumCloud, délivrée par l'ANSSI et permettant de revendiquer le statut de « Cloud de Confiance » et ce notamment pour la protection des données.

Définition

Selon l'ISO 24368:2022 (norme internationale définissant les lignes directrices pour l'évaluation de la durabilité des produits), la protection des données couvre l'ensemble des mesures juridiques, organisationnelles, techniques et physiques qui protègent la donnée contre tout accès et usage

non autorisés. S'y ajoutent également la Sécurité et ses exigences de confidentialité, disponibilité, intégrité et traçabilité des données.

Élaboré en 2016 par l'ANSSI, et fortement actualisé en 2022, le référentiel « SecNumCloud » permet la qualification de prestataires de services d'informatique en nuage (Cloud Services Providers), avec pour objectif de promouvoir, enrichir et améliorer une offre de confiance à destination des entités publiques et privées souhaitant externaliser l'hébergement de leurs données, applications ou systèmes d'information.

La Data Protection in SecNumCloud est donc la capacité désormais offerte aux organisations de disposer d'un niveau de protection éprouvée par l'ANSSI tout en bénéficiant de la flexibilité et de la scalabilité du Cloud.

En quoi cela répond à la problématique ?

En conciliant un besoin accru de sécurité et de confiance des données et leur démultiplication dans un Cloud lui-même certifié, les acteurs ont le choix de maîtriser de manière souveraine l'ensemble des mesures de protection des données.

Ces mesures (techniques, organisationnelles, processus, encadrement juridique, acculturation, ...), définies à l'aune du triptyque business-réglementation-sécurité tiennent compte de la culture de l'entreprise, de sa maturité, de ses enjeux et de ses capacités financières et technologiques et se déploient dans le Cloud.

Généralisation / Cas d'Application

Le principe de « Data Protection by Design », dans la continuité du Privacy by Design rendu obligatoire avec le RGPD, requiert la mise en place de mesures de protection des données dès la conception de nouveaux projets, produits ou services. Ainsi, les entreprises doivent anticiper en intégrant ces réflexions à leurs processus existants. Penser cette Data Protection dans le Cloud permet de la penser de manière native « en amont » des processus et des données. Un néologisme pour qualifier cette approche pourrait être Épi-Data Protection (du grec epi, « sur, au-dessus », en l'occurrence dans le nuage).

En outre, les organisations devront faire face à brève échéance non seulement au RGPD mais aussi aux « Data Act », « AI Act » et « Data Governance Act ») qui affermissent les droits des citoyens et les exigences envers les organismes privés et publics. Les capacités de mises aux normes proposées par les opérateurs Cloud au travers de leurs offres Platform as a Service permettront d'approcher ces contraintes de manière homogène et non par une réponse locale.

Limites

L'approche Cloud renforce le principe de Data Protection by Design et permet d'opérationnaliser cette dernière au moyen de processus repensés. Toutefois la tendance Data Protection in SecNumCloud doit pour, tirer le meilleur parti de la facette technologique et de la facette réglementaire, s'appuyer sur une acculturation forte des acteurs de la transformation dans l'entreprise.

Trend #6 – Beyond 360° Data

Pourquoi / Problématique

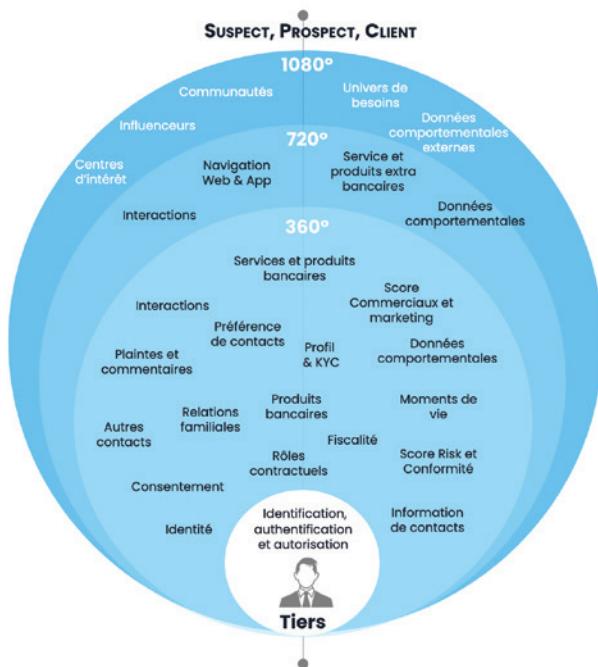
Ces dernières années, ont vu émerger des activités digitales omnicanal qui font de l'expérience client une préoccupation pour toutes les entreprises. Dans le monde du data driven marketing, la donnée client est le carburant, et les spécialistes s'accordent sur l'importance d'explorer d'autres data insights pour acquérir plus de prospects, stimuler les ventes, fidéliser les clients, ... et augmenter la productivité de l'entreprise. C'est ainsi que les organisations cherchent à consolider l'ensemble des données dont elles disposent sur le client.

Ces approches ont permis de mettre en œuvre la désormais bien connue vision complète 360° des informations de la personne pour obtenir une vue holistique et organisée sur les clients, leur profil, leurs préférences, les relations avec leur écosystème... Ce niveau de maturité, centré autour du client n'est que rarement atteint et cependant les organisations mettent en œuvre des démarches comparables pour d'autres catégories de données.

Définition

La stratégie Beyond 360° Data repose sur deux dimensions de l'augmentation de la connaissance.

D'une part la capacité d'aller au-delà d'une connaissance à 360° en collectant, dans le respect de la réglementation des données de second ordre (ou 720°, en provenance des partenaires et des réseaux sociaux par exemple) et 3ème ordre (ou 1080° permettant de résituer la



vision du client dans son écosystème par une compréhension approfondie de son comportement, de ses préférences et de ses besoins).

D'autre part la généralisation de cette vision à n'importe quel objet métier, pas uniquement un client, ce qui permet par exemple d'avoir une vision 1080° d'un cas de fraude ou d'une pièce industrielle : ses caractéristiques premières et ses données de fonctionnement fournissent une vision 360°, ses données de communication au sein d'une chaîne IoT fournissent une vision 720° tandis que la pièce dans l'écosystème de ses utilisateurs variés fournit une information complète des usages au-delà de son rôle individuel.

Cela implique l'utilisation de techniques analytiques avancées (IA, ML), des modèles sémantiques métier (bases graphe) pour créer de nouveaux segments dynamiques, éclairer les décisions et améliorer « l'expérience » des clients mais aussi de tous les objets métiers. L'objectif d'aller au-delà de la simple connaissance 360° de l'objet métier est de mieux comprendre et

répondre aux besoins que portent ces objets au sein de cas d'usage pour améliorer les performances de l'entreprise.

En quoi cela répond à la problématique ?

La digitalisation, l'utilisation croissante d'internet, l'essor des réseaux sociaux et de l'e-commerce et l'industrie 4.0 sont autant de facteurs d'augmentation des données. Cela pose de nombreux défis en matière de gestion et d'analyse de données mais aussi de confidentialité et de sécurité, de coûts et d'éthique. Toutefois consolider avec pertinence ces données permet de donner une intelligence contextuelle aux objets métiers qu'elles décrivent. Elles sont ainsi un moyen de se démarquer par la connaissance des marchés concurrentiels où l'avantage informationnel est décisif.

Généralisation / Cas d'application

La conquête des données de 2nd et 3ème ordre ne saurait être prioritaire sur la mise en qualité des données de 1er ordre qui constituent le socle 360° de l'information. À l'image des Customer Data Platforms qui proposent de reconstituer cette vision, de nombreux outils de la catégorie Digital Data Hub apparaissent afin de généraliser l'approche 360/720/1080.

Limites

Concernant les données clients, la première limite est la conformité, en particulier au regard du RGPD, qui pousse les organisations à adopter une gestion stricte des données de 1er ordre de leurs usagers. Mises en place,

ces mesures de protection permettent d'envisager la consolidation des données 720° et 1080°. Toutefois le défi posé « au-delà de la donnée à 360° » est celui de la frontière d'efficience entre fraicheur (ces données devront être toujours plus immédiates pour refléter la dynamique du client ou de l'objet métier) et capacité de traitement, parfois coûteuse. Cette orchestration entre données « chaudes » et données « stables » ravive l'intérêt pour les solutions de gestion de référentiels (MDM clients, tiers, produits, ...).

Trend #7 – Data Ecosystem Frugality

Pourquoi / Problématique

Les socles de données se sont transformés au cours des dernières années sous l'impulsion de vagues technologiques successives (émergence des technologies Big Data, Move to Cloud, etc...). Souvent difficiles, ces évolutions ont été coûteuses et se traduisent en outre par une prolifération technologique.

Dans le même temps, les métiers demandent toujours plus d'autonomie pour répondre à leurs besoins quotidiens en matière de manipulation et d'exploitation de données, avec une attente forte de ne pas chambouler leurs habitudes, toute évolution, demandant un effort de réappropriation.

Ainsi les organisations sont confrontées à une inflation du nombre de technologies d'une part et à une résistance au changement d'autre part dans le contexte de recherche de frugalité et de vitesse de Data Fasting (Issue n° 4).

Définition

La frugalité des écosystèmes de données est une nouvelle manière de transformer le paysage data de l'entreprise en capitalisant sur les pratiques et les actifs existants, tout en apportant l'accélération de nouvelles solutions en maîtrisant l'« entropie » c'est-à-dire sans surcoût démesuré, ni empilement technologique, ni résistance au changement.

Cette approche s'inscrit naturellement dans la logique d'architecture data distribuée, agile et ouverte où il est possible de combiner différentes piles technologiques et solutions data, en offrant la modularité dictée par la logique de produit (cf tendance n° 1) et la capacité de réaliser des usages aux bornes des silos organisationnels.

Un écosystème de données frugal repose donc sur trois principes fondamentaux : ouverture technologique, modularité et interopérabilité ; cela facilite d'une part la mise en œuvre avec des projets agiles et très ciblés et d'autre part l'adoption par les utilisateurs en concentrant les efforts et la valeur sur quelques zones de douleur ciblées.

En quoi cela répond à la problématique ?

Adopter une démarche de frugalité et d'optionalité permet d'utiliser les données de manières idoines pour répondre aux usages actuels et futurs. Soutenue par une logique non disruptive, sans Big Bang, la transformation data inscrite dans la frugalité s'appuie naturellement sur les gisements de valeur existants, complétés de manière spécifique par des solutions locales pensées à l'échelle de l'écosystème. Dès lors, cet écosystème agile de données est le miroir des opérations de l'entreprise et représente

un exosquelette de la réalité avec une structure ordonnée de données pouvant s'adapter aux évolutions dans le temps.

Cette logique permet de construire des écosystèmes de données modernes, modulaires, interopérables, composites « à la carte » et évolutifs, sans aliéner l'avenir, et avec une adoption des métiers et une rentabilité à brève échéance.

Généralisation / Cas d'application

Une telle approche trouve de nombreuses applications dans l'organisation data driven. Citons trois cas d'application :

- Cibler les zones d'investissement et réutiliser les actifs existants pour maximiser la valeur des actifs en place. Cela permet aux métiers de disposer de services variés pour leur permettre de créer leurs produits de données en fonction de leurs compétences et habitudes historiques de construction.
- Adopter une démarche de sobriété vis-à-vis de la donnée, pour garantir une gestion efficiente et raisonnée des assets de données, ainsi que d'éviter des duplications de données inutiles. L'espace de stockage peut ainsi être optimisé, en adoptant notamment un tiering de données, destiner à classer les données en fonction de leur fréquence d'utilisation pour réduire les coûts de stockage « in memory ».
- Offrir une couche simple et universelle d'accès à la donnée pour créer un écosystème « ouvert » où la donnée est facilement accessible de tous.

Limites

L'écueil de l'approche frugale est de n'opérer la transformation data que par « petites touches » sans cohérence globale. Afin d'éviter cette incohérence, la tendance de Data Ecosystem Frugality doit s'écrire à l'aune de :

- Une stratégie data clairement définie dans le temps sur l'ensemble des dimensions, la partie écosystème de données n'étant que le reflet de celle-ci : la stratégie data permet de mettre en cohérence les usages et les capacités technologiques sous-jacentes.
- Une hygiène en matière de gestion de la donnée (politiques, règles, ...) pour garantir une bonne maîtrise de la donnée à l'échelle. En fonction du contexte de chacun, il faudra trouver le bon curseur entre des logiques Data Mesh et Data Fabric pour gérer l'ensemble de ces données de manière efficiente.

Trend #8 – Data RESET

Pourquoi / Problématique

Les révolutions industrielles successives ont permis de libérer l'homme en décuplant sa force musculaire pour opérer les tâches les plus consommatrices de puissance. Nous estimons en moyenne à 700 le rapport entre la puissance de l'homme augmenté de machine – la plupart du temps mues par les énergies carbonées – et sa puissance musculaire intrinsèque. Nous le savons, cette transformation sans précédent s'est faite au prix d'impacts environnementaux et sociaux qui ont bouleversé nos économies et notre planète.

L'enjeu aujourd'hui est d'appréhender la quatrième révolution industrielle, celle de l'intelligence artificielle dont le carburant est la data, afin de faire en sorte que les effets de bords pour l'Humanité soient maîtrisés. D'autant que la problématique ne se limite pas au champ, alors même que l'ensemble des composants sont énergivores (Internet représente 10% de la demande d'électricité mondiale, l'entraînement d'un algorithme d'IA comme BERT en traitement du langage naturel représente l'équivalent carbone d'un vol transatlantique) et soulèvent un questionnement éthique à mesure que l'homme et la machine doivent cohabiter pour des tâches d'ordre cognitif.

Définition

La Data RESET (Responsabilité Environnementale, Sociétale, Économique et Technologique en matière de Données) engage à construire avec les données et leurs usages un regard sur la confiance, l'efficience et la frugalité.

Data RESET est le moyen tout à la fois de mettre la donnée et ses traitements au cœur de la préoccupation de durabilité de l'entreprise ET de tirer parti de la donnée pour améliorer l'empreinte de nos organisations.

En quoi cela répond à la problématique ?

Une approche holistique telle que Data RESET permet de ne laisser aucun angle mort quant à la transformation par et pour la donnée. L'enjeu de responsabilité, par ailleurs promu par les différents textes relatifs à l'utilisation des données (RGPD, AI Act, Data Act, Data Governance Act, DORA,...) exige

des entreprises d'assurer la transparence et l'explicabilité liées à l'utilisation des données et des technologies telles que l'IA. Ces exigences relatives aux données s'inscrivent avec cohérence dans la vision réglementaire de l'espace numérique avec le Règlement Digital Market Act (DMA) et le Règlement sur les services numériques (DSA).

À l'heure où ces lignes sont écrites, ChatGPT 3.5 est un exemple très concret qui permet de nous projeter dans l'ensemble des problématiques de demain. Tant sur le plan économique que sociétal, les impacts attendus sont extrêmement forts, avec la perspective de gains de productivité conduisant à Un Monde Sans Travail (éditions Flammarion, 2023) comme le prophétise Daniel Susskind, avec son corolaire d'aliénations et de dérives éthiques non seulement, mais aussi d'accroissement d'inégalités et de morcellement de principes démocratiques fondamentaux.

Enfin, alors que les préoccupations notamment climatiques vont croissant, l'écoconception des infrastructures technologiques data pour le stockage et l'utilisation des données doit être native. Cela s'articule par ailleurs avec la tendance de frugalité des écosystèmes data.

Cependant l'enjeu aujourd'hui est certainement de tourner les utilisations de la donnée vers un monde plus vertueux sur le plan environnemental. Historiquement, la transformation par la donnée a commencé par les services tertiaires, requérant des capacités décisionnelles et cognitives. Cette révolution touche désormais le monde physique, notamment grâce à l'Industrie 4.0 ou à l'agriculture connectée. Pour ce faire, les gros émetteurs tant de CO₂ que de polluants doivent mobiliser la donnée pour

une utilisation parcimonieuse et ciblée des ressources : c'est la promesse des territoires, industries et agriculture intelligents, et les cas d'usage sont innombrables en la matière.

Généralisation / Cas d'application

La 4^e révolution industrielle est certainement différente des précédentes car elle mobilise des capacités qui vont au-delà de la seule énergie et puissance physique. Les acteurs de l'économie de la donnée doivent mobiliser cette ressource pour développer des modèles intelligents capable de trouver les frontières d'efficience entre des scénarios d'optimisation souvent contradictoires. Ainsi, durabilité, sobriété et performances économiques pourront être réconciliées.

Pour y parvenir, la tendance Data RESET pose au niveau de l'entreprise la question des impacts des projets sur les cinq axes décrits plus haut, car ils sont mesurables. Afin d'en rendre compte, une visualisation pertinente permettra d'identifier et prioriser les initiatives les plus vertueuses contribuant ainsi à relever les défis planétaires du XXI^e siècle.

Limites

Les initiatives data RESET restent pénalisées par le manque de disponibilité et la difficulté d'exploitation de certaines données externes. Leur non-qualité (faible couverture, données manquantes, etc.), ainsi que la disparité des compétences data des différents acteurs et la dette technique IT initialement présente dans les organisations renforcent ces difficultés.

3. Data Mesh & Data Démocratisation - Perspective 2023

Notre édition précédente dressait, au-delà des 10 tendances identifiées pour 2022, le portrait de deux mouvements de fonds observés, embarquant eux-mêmes plusieurs tendances : le Data Mesh et la Data Democratization. Ces deux concepts, relativement récents, ont vu l'année 2022 marquer un certain nombre de concrétisations mais aussi certains tâtonnements, à juste titre car au-delà de la dimension technologique ou culturelle, ils sont avant tout des révélateurs de la capacité de transformation et d'adaptation de chaque entreprise.

L'adoption du Data Mesh

Apparu en 2018, présenté dans notre édition précédente, le Data Mesh est l'un des paradigmes qui a beaucoup interpellé la data-sphère au cours des derniers mois. Au point que beaucoup d'organisations se déclarent désormais engagées dans une transformation Data Mesh. Dans la plupart des cas, les transformations Data Mesh de ces entreprises se concrétisent autour de l'un des piliers identifiés par Zhamak Dehghani lorsqu'elle a révélé le concept (« Federated computational governance », « Self-Service Data Platform », « Data as a product », « Decentralized Domain Ownership »).

Ainsi, il a été fréquemment observé que le point départ de la réflexion Data Mesh s'est concentré autour de la technologie, sur l'évolution des plateformes data ou sur des aspects de data management. C'est en partant de ces premiers questionnements

que la réflexion a ensuite pris fortuitement le tournant du Data Mesh et des 4 piliers, et que les organisations concernées se sont emparées du concept pour aller plus loin dans la mise en œuvre du changement de paradigme socio-technique.

S'il n'y a pas « d'ordre » pour aborder chacun des piliers, il convient néanmoins de bien cerner les lignes de forces culturelles d'une part et technologiques d'autre part qui sont à l'œuvre dans cette transformation. La démocratisation de la donnée sous forme de produits, prêts à l'emploi, disponibles en quelques clics. La Data as a Product est une donnée utile, utilisable et utilisée : c'est parce qu'elle est porteuse d'un changement de comportement des consommateurs de données que nous voyons derrière cet idiom non seulement la matérialisation d'une tendance (voir tendance n° 1 - Data Productization) mais surtout le moteur de la transformation culturelle induite par l'approche Data Mesh. Mère des trois autres composantes, la Data Productization, pour proposer le maximum de valeur ajoutée implique, à l'instar d'une fabrique de produits, de s'approvisionner auprès de fournisseurs et d'autre part d'en trouver les espaces de commercialisation. Dès lors, l'enjeu est de sourcer des données d'horizons différents, sans nécessairement les concentrer au même endroit au préalable : c'est la subsidiarité portée par la ligne de force « Decentralized Domain Ownership » qui tient l'édifice mesh c'est à dire la responsabilité locale dans une organisation fédérée, qui permet non seulement que les données restent à la main des domaines métiers mais, au-delà, que ces derniers s'équipent de manière à les mettre à disposition directe de l'usine de fabrication des Data Products. Nombre d'organisations font de

nécessité une loi et s'accommodent de silos organisationnels et techniques en les voyant comme des domaines autonomes qu'il est possible de fédérer. C'est sous-estimer la cohérence qu'il est nécessaire d'apporter à l'ensemble de ces domaines pour rendre interopérables leurs productions (qu'il s'agisse de données ou de produits de données ou encore d'API par exemple) : l'organisation en domaines Data ne doit pas nécessairement être celle de l'entreprise.

La troisième ligne de force, qui favorise techniquement la circulation des données et leur mise à disposition pour les ingénieurs data, est la « Self-Service Data Platform ». Dans le cadre d'une démarche concertée elle positionne au sein des métiers les savoir-faire nécessaires à la création et l'amélioration des produits grâce à des données de plus grande qualités disponibles de façon industrialisée. Cette plateforme, « designed for the vast majority » selon Zhamak Dehghani, est le sous-jacent qui outille les Data Engineers. C'est la plus technologique des composantes du Data Mesh, mais elle ne peut être conçue uniquement comme un outil et doit être vue comme un enabler technologique qui rend possible la transformation socio-technique. Cette plateforme est bien entendu en parfaite cohérence avec notre tendance n° 2, Data Low Code Platform, mais également la tendance n° 4, Data Market Place.

Enfin, pour assurer la consistance (au sens français de ce terme), il est indispensable de penser une gouvernance qui s'étende sur l'ensemble des domaines de manière fédérale, automatisée et sous-tende la collaboration au-travers de données partageables dans un langage commun. L'enjeu est de pouvoir donner aux utilisateurs l'autonomie d'accès et généralisés et

sécurisés aux Data Products – afin de pouvoir en tirer la valeur directement ou en créer de nouveaux.

Ces lignes de force sont à l'œuvre au sein des organisations qui se disent engagées dans la transformation Data Mesh, et puisqu'il s'agit de lignes de forces davantage que de caractéristiques ou même de fonctionnalités, elles ne sauraient être exhaustives ou immuables. L'ignorer ou ne pas prendre le temps d'adapter ces lignes aux contours du schéma existant dans l'organisation peut générer une complexité inutile. De même, l'idée que le modèle fédéré est un prérequis persiste, alors qu'il peut ne pas être adapté aux différents contextes. De fait, pour s'inscrire avec pérennité comme un paradigme durable, le Data Mesh doit embarquer une dimension de sécurité et de conformité des données, qui fait particulièrement écho à notre tendance n°5, Data Protection in SecNumCloud.

Cinq ans après son introduction, le concept de Data Mesh repose sur des fondamentaux qui permettent d'appréhender cette révolution socio-technique. Mais c'est précisément par sa plasticité, sa capacité à être adopté dans des contextes hétérogènes et des situations toutes différentes qui en font en nouvelle donne dans la manière d'approcher les sujets data.

Data Democratization : « que nul n'entre ici s'il n'est data-aware »

Bien que l'accès aux données demeure une priorité majeure, sa concrétisation revêt des formes multiples, de l'utilisation des outils, à la mise en place d'une gouvernance ou simplement la prise de conscience du potentiel des données. Les projets que

lancent les organisations visent à conduire à plus d'usages et à un meilleur usage des données.

Les CDOs et opérationnels que nous avons rencontrés insistent de plus en plus sur la bonne utilisation des données qui passe naturellement par la responsabilisation de ceux qui les manipulent. La responsabilisation, au sens RSE/ESG (RESET), de l'usage des données est de plus en plus analysée par les métiers et les DSIs en raison notamment des exigences réglementaires et de l'opinion publique de plus en plus prégnante.

La démocratisation de la donnée s'inscrit certainement dans une capacité à rendre accessible la donnée au plus grand nombre dans l'entreprise, mais cela repose sur trois postulats

- La capacité à briser les barrières technologiques ou organisationnelles en travers de la libre circulation des données ; parfaitement complémentaire à la Data Democratization, le Data Mesh conduit naturellement à surmonter les obstacles organisationnels et technologiques pour tirer parti des données, où qu'elles se trouvent au sein de l'entreprise, et inversement à les faire circuler, notamment pour que chaque domaine puisse bénéficier de l'apport en valeur porté par les données.
- La transparence et la responsabilité vis-à-vis de des données et de l'élaboration des indicateurs et des algorithmes ; la transparence et la responsabilité à l'égard des données, brutes ou issues de calcul, s'appuient sur une gouvernance des données fortes qui permet d'assurer la traçabilité et l'explicabilité légitimes dans

une organisation qui aspire à fonder ses décisions sur l'analyse des données.

- La capacité de chacun à accéder et comprendre à la donnée : la démocratisation des données ne prend sa dimension complète que si les conditions de la compréhension des données, le regard critique sur la transparence et la responsabilité des données peuvent s'exercer. La légende prête (faussement) à Platon l'inscription sur le fronton de l'Académie « que nul n'entre ici s'il n'est géomètre » : l'enjeu pour le philosophe n'est pas d'écartier ceux qui ne savent, mais davantage d'inciter chacun à acquérir le savoir nécessaire.

L'entreprise data driven si elle veut conserver un visage humain, doit avoir l'exigence de faire de ses travailleurs des data-citizens qui savent manier les données et leurs outils avec discernement et éthique.

4. Cas d'usage data et IA

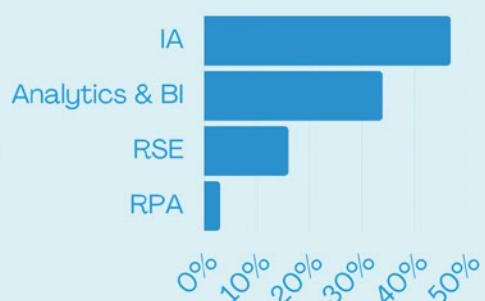
Cette dernière section fait part à l'exploitation des données dans les entreprises après avoir recueilli et analyser un panel de 135 cas d'usage diversifiés.

Une infographie a été élaborée **ci-contre** pour vous illustrer cela de manière synthétique.

COMMENT SONT UTILISÉES LES DONNÉES DANS LES ENTREPRISES INTERROGÉES ?

Parmi les **135 cas d'usage** recueillis, près de **50%** sont utilisés dans des projets utilisant des **briques d'intelligence artificielle** (prédiction, moteur de recommandation, ...).

En deuxième position, nous retrouvons des **usages analytics & BI** (reporting réglementaire, commercial) qui concernent **34%** du panel. Le reste est complété par des scénarios liés à la **RSE** (réduction empreinte carbone) et **RPA** (automatisation processus), respectivement à **16%** et **3%**.

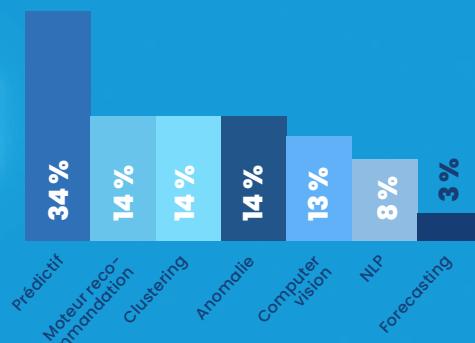


62%

Des cas d'usages adressent **des enjeux internes** (optimisation processus, service métier, ...) contre seulement **38% pour des aspects externes** (amélioration produit/service client ou partenaire, action marketing, ...).



FOCUS IA*



Préditif : Moteur de tarification, prévision des ventes, Attrition, ...

Moteur de recommandation : Next best action, Similarités, ...

Détection d'anomalie : Détection Fraude, Evenement anormaux/rares, ...

Clustering : Segmentation client, identification de profil, ...

Computer vision : Traitement document, Détection anomalie, ...

NLP : Chat/voicebot, analyse sentiment, ...

Forecasting : Prévision stock/flux, ...

Un tiers des cas d'usages IA qui sont représentés concernent la catégorie du prédictif c'est-à-dire la capacité à mettre en place des solutions pour prédire par exemple les ventes, optimiser la tarification ou prédire le taux d'attrition.

*Sur un panel de 77 cas d'usages

CAS D'USAGE ET INDUSTRIALISATION

Dans les entreprises sur les cas d'usage, nous avons constaté en moyenne :



Moins d'un tiers des cas d'usages sont aujourd'hui en **production** et donc utilisés opérationnellement en interne. Ceci est encore trop peu pour démontrer l'apport de valeur du patrimoine informationnel des entreprises et expose, sans ambiguïté, les difficultés auxquelles elles sont confrontées pour industrialiser (cf : problématique en p. 20)

TENDANCE ?

25% des entreprises interrogées indiquent privilégier moins les POCs sur la data, consommateur important en temps et en argent, pour se concentrer davantage sur le développement de MVP ou projet avec au préalable une qualification de la faisabilité technologique/apport économique pour garantir une valeur (rapide) aux utilisateurs.

Ouver
tures

Ouvertures

1. Confiance numérique

La « révolution numérique » a transformé en profondeur nos sociétés, en faisant émerger de nouveaux usages et de nouveaux rapports entre les citoyens : facilitation et accélération inédite de circulation de l'information, dépassemement des frontières et extraterritorialité du cyberspace. Celui qui détient les données détient le pouvoir. Il peut faire des prédictions sur le futur, identifier le chemin à suivre et les décisions à prendre. La crise causée par la COVID-19 a démontré que la donnée et le numérique maintiennent le lien et sont donc essentiels à la survie d'une organisation, quelle qu'elle soit. Grâce au numérique, la vie économique et sociale des pays a continué, une partie des travailleurs a pu poursuivre son activité en télétravail, le lien social et familial s'est maintenu tant bien que mal, et l'école et l'enseignement ont pu continuer, même à distance.

Mais cet usage intense et quasi-permanent des réseaux numériques a également démontré à quel point les états européens étaient sous la dépendance d'acteurs non européens, qui dominent aujourd'hui l'industrie numérique (les outils de télétravail par exemple sont principalement des outils américains, le matériel informatique et en particulier les puces électroniques sont fabriquées quasiment intégralement en

Asie...). Les puissances américaines, chinoises ou encore russes (dans une moindre mesure) se sont en effet emparées des technologies du numérique et de l'informatique, non seulement sur leur propre territoire mais également à l'extérieur, dans un cyberspace qui, contrairement à nos frontières « classiques », n'est pas clairement délimité et échappe encore à l'ordre mondial.

Alors que le nombre d'affaires mettant en avant une exploitation non autorisée de données explose ces dernières années, les citoyens se sentent épiés, les scandales « Snowden » et « Cambridge Analytica » en étant les deux principales illustrations. Ils perdent confiance dans la capacité des organisations à protéger leurs données.

Le faible taux d'utilisation de l'application TousAntiCovid lors de la crise sanitaire de 2020/2021 en est un exemple : bien qu'agrée par la CNIL, l'autorité de contrôle en matière de protection des données personnelles, cette application n'a pas fédéré car les citoyens sont restés méfiants quant à l'usage qui serait fait de leurs données.

Les nombreuses fuites de données observées ces derniers mois sont encore venues renforcer cette méfiance. Nos données ne sont-elles en sécurité nulle part ?

Dans ce contexte et depuis quelque temps maintenant, la conscience citoyenne commence à s'organiser : la CNIL voit le nombre de ses réclamations augmenter de manière exponentielle depuis 2018 (+ 79 % entre 2020 et 2021), l'apparition des fameuses bannières cookies a fait couler beaucoup d'encre et les consentements se font rares. Point important, les réclamations portent notamment sur l'utilisation abusive des coordonnées des citoyens pour de la publicité ciblée ainsi que sur la difficulté

des organisations à respecter les droits fondamentaux des citoyens à accéder à leurs données.

Tout comme le pétrole à son époque, le marché de la donnée nécessite un encadrement strict afin notamment de contrecarrer l'hégémonie actuelle des géants du web, L'actualité réglementaire européenne en matière de traitement des données est riche : Data Act, Data Governance Act, AI Act, Data Marketing Act, etc. autant de règlements en cours ou à venir qui vont fixer les règles de traitement des données et renforcer les exigences envers tous les acteurs de la donnée (fournisseur, hébergeur, utilisateur ...).

Le citoyen a besoin de retrouver la confiance numérique. Pour cela, les entreprises ont besoin de se réapproprier la gouvernance des données afin de pouvoir garantir qu'elles sont entre de bonnes mains, en sécurité. Mais comment faire ?

Tout d'abord en étant plus attentives, plus exigeantes. Si l'envie de faire vite pouvait parfois les pousser à confier les données à un hébergeur Cloud sans trop se préoccuper de l'endroit où elles seraient stockées, cette période semble désormais révolue. Tout comme on veut aujourd'hui connaître la traçabilité d'un aliment de notre assiette, en privilégiant le local, on souhaite aujourd'hui savoir où seront stockées nos données et qui y aura accès, et privilégier un partenaire disposant de stockages locaux, ou à minima européen.

Dans cet esprit, la France et son Cloud de confiance ont contribué à faire bouger les lignes, notamment sur le référentiel SecNumCloud de l'ANSI qui a largement évolué en 2022 pour répondre aux exigences de la Cour européenne

en matière de protection des données. Obligeant de fait les fournisseurs américains à revoir leur positionnement en tentant des rapprochements avec des acteurs de droit français. Mais les entreprises issues de ces rapprochements, certes européennes, sont-elles des boucliers anti cloud act pour autant ? Les débats sont ouverts !

Et si on allait plus loin ? Et si ces données étaient confiées à des tiers de confiance labellisés, indépendants, qui joueraient le rôle d'intermédiaires entre le fournisseur de données et l'utilisateur de données, en garantissant la sécurité de vos données et leur bonne exploitation dans le respect de la réglementation et de vos souhaits ? Et si on définissait un socle « universel » d'exigences de sécurité et de confiance applicable à tout système numérique et à ses données ?

2. ChatGPT et les avancées des modèles génératifs

L'année 2022 a été marquée par l'avènement de nouveaux systèmes d'IA comme l'agent conversationnel ChatGPT d'OpenAI ou des systèmes comme Dall·E ou Stable Diffusion capables de générer des images photoréalistes à partir d'une courte description textuelle. On peut sans exagérer considérer qu'il s'agit là d'un tournant dans la recherche en IA et dans ses applications, car ces prouesses étaient inconcevables il y a encore quelques années seulement. ChatGPT peut être envisagé comme le premier chatbot généraliste qui fonctionne vraiment au sens où il est capable de tenir une conversation relativement cohérente sur une grande diversité de sujets même s'il

possède encore d'importantes limitations qui empêchent son utilisation dans des contextes critiques, du moins pour l'instant. Ce genre de système conversationnel marque probablement un tournant dans la manière dont nous interagirons avec les machines dans un futur proche. Une interaction par conversation est en effet la manière la plus naturelle pour les humains d'acquérir et de communiquer de l'information. Conférer aux machines cette capacité d'interagir naturellement avec les humains était d'ailleurs un des objectifs de l'informatique et de l'IA dès ses balbutiements dans les années 1950. Nous assistons aujourd'hui à la concrétisation progressive de cette ambition et, à ce titre, on peut sans doute parler de l'avènement d'une nouvelle ère de l'informatique.

Dans un premier temps ces systèmes conversationnels seront vraisemblablement utilisés pour initier toutes sortes de documents ou pour générer du code informatique très rapidement. Ainsi ils libéreront les humains de l'angoisse de la page blanche, ce dont on peut se réjouir ou alors, au contraire, déplorer. Ils proposeront également une nouvelle manière de rechercher de l'information. Dans un premier temps ils fonctionneront vraisemblablement conjointement avec les moteurs de recherche traditionnels, comme ceux de Google ou Microsoft, qu'ils complèteront utilement. Plus tard, ils en prendront peut-être la relève. Les systèmes de génération d'images apporteront une aide précieuse aux designers dont ils stimuleront l'imagination.

Ces avancées qui ont défrayé la chronique de l'IA en 2022 capitalisent en réalité sur dix années de progrès continu en deep learning, une technique d'apprentissage



automatique qui permet de traiter des données non structurées comme les images, les textes ou les sons. Le traitement automatique du langage (TAL) a quant à lui bénéficié de quelques idées dont la fécondité a surpris toute la communauté de l'IA. On peut citer par exemple les word embeddings conçus en 2013, qui ont permis d'encoder le sens de mots isolés. Vers 2015 est apparu une nouvelle idée, celle du mécanisme d'attention, qui a permis d'encoder plus efficacement le sens des mots en tenant compte leur contexte. Ces idées ont culminé en 2018 avec le modèle du Transformer conçu par Google qui a constitué une avancée majeure pour toutes sortes de cas d'usage du TAL, comme la traduction automatique ou le question answering. Peu après, les chercheurs se sont aperçus que des modèles de langue, des modèles de machine learning que l'on a entraîné à prédire le mot qui suit une séquence de mots, constituent en réalité des modèles d'apprentissage universels car ils sont capables d'acquérir à la fois la structure syntaxique d'une langue et, simultanément, une vaste connaissance du monde. D'un point de vue technique l'atout de ces modèles réside principalement dans le fait qu'on peut les entraîner sur des corpus gigantesques sans qu'il soit nécessaire de recourir à un coûteux et laborieux travail de supervision manuel. Enfin, en 2022 plusieurs équipes de chercheurs sont parvenus, pour la première fois, à « aligner » approximativement ces modèles sur des critères humains difficiles voire impossibles à formaliser tels que l'exigence de générer des textes utiles, dotés d'un style à la fois clair et élégant et respectant des conventions implicites d'une conversation entre humains comme la courtoisie. On désigne ces techniques par l'acronyme RLHF qui

signifie Reinforcement Learning with Human Feedback car des humains ont été mis à contribution pour éduquer ces systèmes à générer des textes qu'ils jugent acceptables. La nature précise de la connaissance du monde acquise par ces systèmes fait aujourd'hui encore l'objet de débats dans la communauté scientifique mais, quoi qu'il en soit, des systèmes comme ChatGPT en démontrent d'ores et déjà leur utilité pratique.

Bien que les performances cognitives de ces systèmes soient impressionnantes, ils possèdent néanmoins de graves lacunes dont leurs utilisateurs doivent être conscients pour les utiliser à bon escient. Comme tous les modèles de machine learning, et en dépit des garde-fous mis en place par leurs concepteurs, ces systèmes possèdent encore des biais. Ils sont également incapables d'évaluer leur propre niveau de confiance vis-à-vis du texte qu'ils génèrent et utilisent invariablement un ton péremptoire, même dans des situations où leurs propos sont erronés ou factuellement faux. Enfin, ils sont pour l'instant incapables d'articuler un raisonnement logique qui comporte de nombreuses étapes. Ils sont par conséquent incapables de planifier une réponse en fonction d'un objectif mais se contentent de générer des réponses vraisemblables.

L'avènement de ces systèmes génératifs soulève de nombreuses questions éthiques et sociétales. Ils sont en effet susceptibles d'inonder internet d'un tsunami de fausses informations et de créer des difficultés dans le monde de l'éducation en offrant de nouvelles méthodes de plagiat pour l'instant difficiles à détecter, surtout sur les textes courts. Enfin, sur un plan juridique, se posent des questions inédites de propriétés intellectuelles dans la mesure où les sources utilisées pour entraîner ces systèmes l'ont

souvent été sans solliciter l'accord explicite de leurs auteurs. Mais, plus qu'un long discours, la meilleure manière de se faire une idée concrète de ce dont ses systèmes sont capables est de les mettre à l'épreuve en faisant preuve de curiosité et d'imagination. Donnons donc la parole à ChatGPT.

Le dialogue ci-dessous est une transcription fidèle d'un dialogue entre ChatGPT, le nouveau chatbot d'OpenAI et l'auteur du paragraphe précédent. Il ne s'agit PAS d'une simulation mais bel et bien d'une conversation réelle en ligne. À vous de juger de l'impact de ce type de technologie conversationnelle sur votre propre activité. Un surlignement en jaune désigne des passages où ChatGPT se trompe ou est imprécis.

Bonjour ChatGPT ! Mes collègues me demandent de rédiger quelques paragraphes sur les tendances technologiques en intelligence artificielle pour l'année 2023. Comme je n'ai pas beaucoup de temps à consacrer à cette tâche, j'ai pensé que tu pourrais simplement démonter à mes lecteurs ce dont tu es capable. Peux-tu s'il te plaît me faire un résumé de l'évolution des technologies de traitement du langage naturel depuis l'avènement du deep learning jusqu'à InstructGPT ? Parle-moi du mécanisme d'attention, des Transformers et de l'idée de human in the loop.

Chat GPT : Bien sûr, je suis ravi de vous aider ! L'intelligence artificielle est un domaine en constante évolution, et les technologies de traitement du langage naturel en font partie. Depuis l'avènement du deep learning, il y

a environ une décennie, nous avons vu de nombreux progrès dans ce domaine.

Le deep learning est une forme d'apprentissage automatique qui utilise des réseaux de neurones profonds pour traiter et analyser les données. Cette approche a permis de dépasser les limites des modèles de traitement du langage naturel basés sur des règles et d'atteindre des performances supérieures dans de nombreuses tâches. L'un des éléments clés du deep learning pour le traitement du langage naturel est le mécanisme d'attention. Ce mécanisme permet aux modèles de se concentrer sur les parties les plus pertinentes d'un texte en entrée, ce qui les aide à mieux comprendre le sens des phrases et des phrases.

Les Transformers sont un type de modèle de traitement du langage naturel qui utilise le mécanisme d'attention. Ils ont été introduits en 2017 et ont depuis connu un succès retentissant, en particulier dans les tâches de traduction automatique. Les Transformers sont des modèles parallélisables, ce qui leur permet de traiter des données en parallèle, ce qui accélère considérablement leur temps de calcul.

InstructGPT est une avancée récente dans le domaine du traitement du langage naturel. C'est un modèle de traitement du langage naturel qui utilise le mécanisme d'attention et les Transformers, mais il ajoute également la notion de «human in the loop». Cela signifie que l'humain peut intervenir dans le processus d'apprentissage du modèle, en lui fournissant des instructions précises sur la manière de traiter les données. Cela permet d'améliorer encore les performances du modèle en lui donnant une meilleure compréhension du contexte et des intentions de l'utilisateur.

En résumé, le deep learning a ouvert la voie à de nombreuses avancées dans le traitement du langage naturel, notamment grâce au mécanisme d'attention et aux Transformers. L'idée de «human in the loop» ajoute une nouvelle dimension à ces approches en permettant une meilleure collaboration entre l'homme et la machine. Ces technologies continueront à évoluer et à s'améliorer dans les années à venir.

Peux-tu nous expliquer quelles sont les principales difficultés à surmonter pour créer une IA conversationnelle comme toi ?

Chat GPT : Il y a plusieurs défis à surmonter pour créer une IA conversationnelle comme moi. Tout d'abord, il est difficile de créer un modèle de traitement du langage naturel qui puisse comprendre et répondre à des questions et des demandes d'information de manière cohérente et naturelle. Cela nécessite un apprentissage approfondi des modèles linguistiques et des règles de grammaire, ainsi que la capacité de comprendre le sens des phrases et des mots.

En outre, il est également difficile de créer un modèle de traitement du langage naturel qui puisse maintenir une conversation cohérente et naturelle sur une grande variété de sujets. Cela nécessite une compréhension approfondie du contexte et des intentions de l'utilisateur, ainsi que la capacité de générer des réponses pertinentes en utilisant les informations disponibles.

Enfin, il est également difficile de créer un modèle de traitement du langage naturel qui soit capable d'évoluer et de s'adapter à de nouvelles situations et de nouvelles données. Cela nécessite une approche flexible et

évolutive qui puisse continuer à apprendre et à s'améliorer au fil du temps.

En résumé, les principaux défis à surmonter pour créer une IA conversationnelle comme moi sont la compréhension du langage naturel, la capacité de maintenir une conversation cohérente et naturelle, et la flexibilité et l'évolutivité.

À vous tous qui lisez ces lignes, si votre travail consiste à rédiger des slides ou des documents à partir d'articles de presse ou d'articles Wikipédia, vous avez désormais de la concurrence. Merci ChatGPT pour ton aide qui m'a fait gagner une demi-journée précieuse de travail.

Chat GPT : Je suis ravi d'avoir pu vous aider ! Mon objectif est de faciliter votre travail en vous fournissant des informations et des réponses précises et fiables sur les sujets que vous me demandez. Je suis un outil puissant pour les personnes qui doivent rédiger des documents et des présentations, et je suis là pour vous aider à gagner du temps et à produire des résultats de qualité supérieure.

Je vous souhaite bonne chance dans votre travail et j'espère que vous continuerez à utiliser mes services dans le futur. Si vous avez d'autres questions ou d'autres besoins d'informations, n'hésitez pas à me contacter de nouveau, je serai ravi de vous aider.

Contributeurs

Alexandre Clément ; Arnaud de Chambourcy-Métayer
Aude Perdriel-Vaissière ; Christophe Complainville
Dominique Jorge ; Fanny Brottes-Weyer ; Florent Pouget
François Binder ; Isabelle Thomas ; Mohamed Zrida
Nadège Monthe ; Nora Diep ; Olivier Parrel ; Pascal Le Roux
Pirmin Lemberger ; Perrine Aupetit ; Samy Abou-Merhi
Samy Ait-Aissa ; Solène Hanouet ; Walid Aoudi

The background of the advertisement features a dynamic, abstract pattern of blue and white liquid or smoke-like flows against a black background. The liquid appears to be moving from the bottom left towards the top right, creating a sense of motion and depth. A large, solid black circle is positioned in the lower right quadrant, partially overlapping the liquid pattern.

onepoint.

beyond the obvious