

N°4 | BAROMÈTRE

DES DIRECTIONS DATA 2022

Plus de 50 dirigeants
s'expriment sur leur trajectoire
de maturité data et IA



Quantmetry

Executive summary

Pour la 4^{ème} année consécutive, Quantmetry a mené une étude qualitative auprès d'une cinquantaine de dirigeants et responsables data de grandes et moyennes entreprises françaises, pour analyser leurs modèles d'organisation, leurs projets, leurs enjeux et leurs difficultés.

Le premier constat de notre baromètre 2022 est la normalisation tant attendue du taux d'industrialisation des cas d'usage IA qui bondit ainsi en un an de 27% à 46%. On note l'atteinte d'un seuil optimal pour une part non négligeable des entreprises ayant investi dans leur transformation data depuis plusieurs années : en moyenne, après deux à quatre années d'investissements et d'expérimentations, les entreprises franchissent la barre des 50% de cas d'usage industrialisés.

Le déploiement progressif des pratiques de MLOps dans plus des 2/3 des entreprises, les investissements pour certains dans des plateformes Low Code, et surtout la montée en maturité des organisations et des compétences expliquent cette hausse.

Mais une fois le sommet de l'industrialisation des cas d'usage IA atteint, une autre montagne se dresse devant les entreprises : la bonne adoption et performance des usages en phase d'exploitation. Même si le MLOPs se généralise petit à petit et fluidifie la relation IT-data, la collaboration avec les métiers dans la gestion du run est quasi inexistante. Ainsi, l'enjeu n'est plus à l'industrialisation technique mais à l'adoption humaine des usages data & IA.

Si la valeur générée en production n'est que trop rarement mesurée, les directeurs data lancent des cas d'usage considérés comme étant de plus en plus créateurs de valeur. Les investissements sont rentabilisés via une variété de leviers économiques, offrant principalement des gains d'efficacité opérationnelle (62% des usages) ou du marketing & des ventes (38% des usages). Ces résultats prometteurs sont majoritairement la conséquence d'investissements dans les fondamentaux de la transformation data : mise à disposition de données riches et de qualité d'une part, investissement dans les infrastructures et les plateformes d'autre part.

L'époque n'est plus aux cas d'usage IA spectaculaires mais à la concentration sur des cas d'usage, souvent plus simples, mais créateurs de valeur. Quelques entreprises déploient des projets IA dont l'envergure et l'impact économique sont élevés. Mais cela reste encore exceptionnel et la majorité du marché est encore loin de capturer tout le potentiel que la data offre dans la transformation des entreprises.

Pour capturer tout ce potentiel, les entreprises devront relever 3 défis majeurs. Tout d'abord, il faudra qu'elles fassent aboutir leur modèle organisationnel. 83% des entreprises matures privilégient un modèle hybride avec un Data Office central et des compétences data au sein des entités Métiers. Ce modèle qui promeut une certaine décentralisation des usages de la donnée est néanmoins complexe à mettre en œuvre. Le déploiement de l'organisation à l'échelle de l'entreprise, la mise en place et l'animation de la gouvernance de la donnée ainsi que l'évolution en cohérence de l'architecture technique restent des défis de tous les jours pour les data officers. Pour y faire face on constate un recours récent mais accru à de nouveaux paradigmes comme le Data Mesh. Si le concept reste encore défini de façon hétérogène selon les entreprises et pris essentiellement par le prisme technologique, il est amené à se développer tant la promesse d'une gestion de la donnée décentralisée mais cohérente est forte.

Le 2nd défi à relever est celui du recrutement et de la fidélisation des compétences data sous grande tension. Cela est particulièrement vrai pour les profils de Data Engineers et Machine Learning Engineers.

Enfin, les entreprises doivent encore combler le fossé culturel entre data et métier. Après avoir travaillé sur le nécessaire rapprochement entre data et IT pour une réussite technologique des projets data, les entreprises doivent désormais réussir à former un trio data, IT et métiers pour que l'adoption des cas d'usage data soit pleinement effective.

Jonathan Cassaigne,
Directeur des Expertises de Quantmetry

Présentation et synthèse du baromètre

08

01

DES RÉSULTATS ÉCONOMIQUES ENCOURAGEANTS, PONCTUÉS DE QUELQUES DÉSILLUSIONS

18

SUCCESS STORIES

#1 Secteur de l'assurance	30
#2 Secteur CPG / retail / luxe	32

02

EN ROUTE VERS LE RUN ! LES ENTREPRISES FRANCHISSENT UN NOUVEAU CAP D'INDUSTRIALISATION DE L'IA

36

DE QUOI PARLE-T-ON?

- Industrialisation et exploitation de l'IA	40
- IA de confiance	52

03

TROIS DÉFIS À RELEVER GÉNÉRER PLUS DE VALEUR

56

DE QUOI PARLE-T-ON?

- Data Mesh	68
- Machine Learning Engineer	74

Conclusion et perspectives

92

Qui sommes-nous ?

100

SOMMAIRE

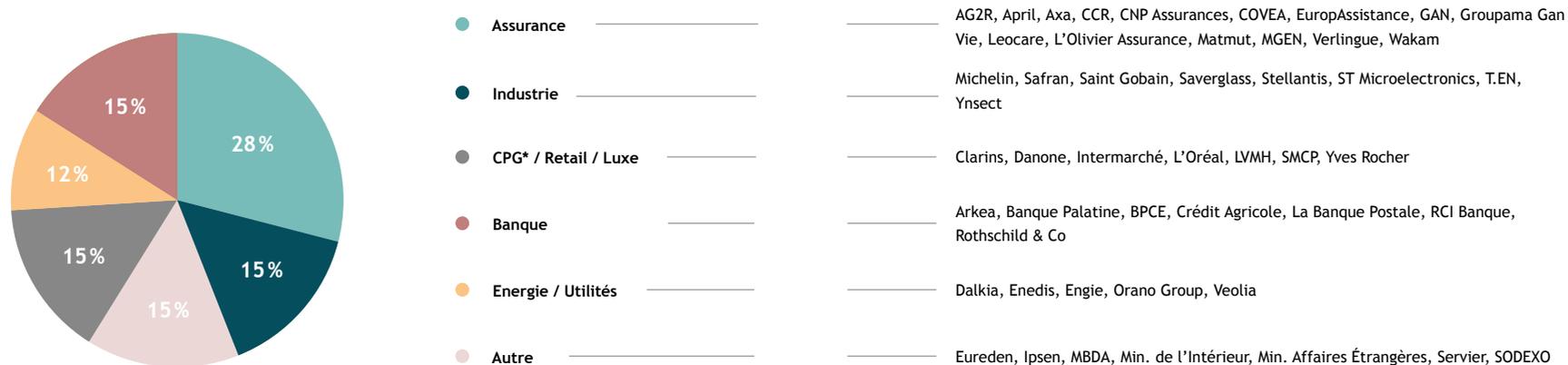
Présentation et synthèse du baromètre

Un panel de 53 répondants couvrant principalement 5 secteurs d'activité

Nous avons interrogé 53 dirigeants et responsables data de grandes et moyennes entreprises françaises. Au cours d'entretiens téléphoniques ou physiques approfondis, nous avons évoqué leur organisation, leurs réussites, leurs challenges et leurs difficultés. L'objectif de ces échanges était de :

- Analyser les progrès réalisés depuis notre Baromètre 2021
- Comprendre comment les directions data abordent leur nouvelle phase de maturité
- Mettre en évidence les enjeux forts partagés par la majorité des directions data

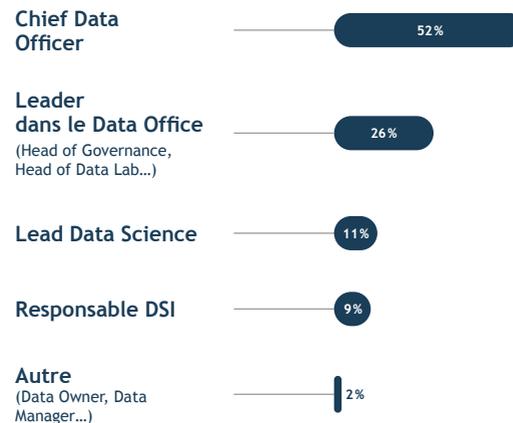
Répartition des répondants par secteurs d'activité



* CPG = Consumer Packaged Goods ou biens de consommation.

Postes occupés

% de répondants



Des résultats encourageants, ponctués de quelques désillusions

- Partie 1 -

- Les investissements continuent de croître et de générer des résultats prometteurs
- Au cours des dernières années, les entreprises ont mené une course aux cas d'usage pour générer de la valeur rapidement. Cette année, beaucoup d'entre elles ont décidé de se recentrer sur les fondamentaux : centralisation, accessibilité et qualité des données
- Le retour aux fondamentaux s'accompagne d'une dépriorisation des projets d'IA. Le marché semble traverser une phase de désenchantement envers l'IA, à l'issue de laquelle les entreprises gagnent en lucidité sur le ratio réel de coût vs bénéfice
- Les entreprises qui persèverent dans l'IA, au-delà des désillusions, y trouvent de puissants leviers de valeur (efficacité du marketing et/ou des opérations, modification du business model, nouvelles sources de chiffre d'affaires...)

77% des sondés estiment que leur budget est bien dimensionné

La part de l'IA dans les cas d'usage stagne autour de

30%

alors que les entreprises repriorisent les projets de décisionnel et de data management

68%

des sondés déclarent que le bénéfice moyen tiré des usages data et IA augmente

62%
de la valeur dégagée par les usages provient de gains d'efficacité...

...contre **38%** issus de CA additionnel

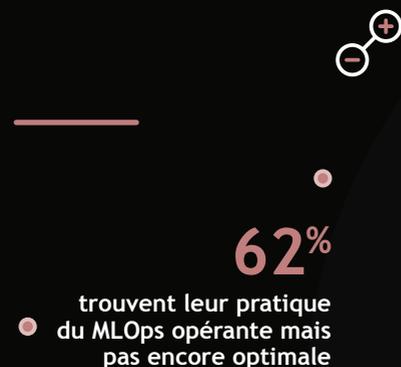
En route vers le run ! Les entreprises franchissent un nouveau cap d'industrialisation de l'IA

- Partie 2 -

- Entre 2020 et 2021, l'IA est véritablement sortie du laboratoire
- Au cours de l'année 2021-22, l'industrialisation de l'IA a commencé à se normaliser
- Les progrès proviennent en grande partie d'améliorations portées sur la stratégie, l'organisation et les socles technologiques
- De nombreuses entreprises accélèrent également leur industrialisation grâce aux outils Low Code / No Code ainsi qu'aux solutions sur étagère
- Le prochain défi : exploiter pleinement les cas d'usage IA en run

La part des entreprises
pratiquant le MLOps est passée de

47% à **72%**



Le taux d'industrialisation moyen
des cas d'usage IA a bondi

de **27%** à

46%

Trois défis à relever pour générer plus de valeur

- Partie 3 -

Défi #1 - Faire aboutir le modèle organisationnel

- À mesure qu'elles mûrissent, les entreprises se tournent vers les modèles organisationnels hybrides offrant de multiples bénéfices
- Mais l'implémentation de ces modèles n'est pas aisée, et elle demeure inaboutie pour près de la moitié des entreprises. Pour aller plus loin, ces entreprises envisagent souvent d'accompagner leur transformation organisationnelle de changements profonds dans l'architecture data

Défi #2 - Sécuriser des compétences sous tension

- Les entreprises sont en manque chronique de compétences data, particulièrement sur les profils techniques comme les Data Engineers
- Les Machine Learning Engineers sont de plus en plus plébiscités pour maîtriser le run des projets d'IA

Défi #3 - Comblent le fossé culturel entre data et métiers

- Les métiers n'intègrent pas encore pleinement la data dans leur pilotage. Trop souvent, la data ne s'inscrit pas dans les grands programmes stratégiques, et les métiers ne se sentent pas responsables du ROI data
- Les directions data quant à elles, sont handicapées par un pilotage trop opérationnel, souvent décorrélé de la valeur délivrée au métier. Sans définition commune de la valeur, data et métiers opèrent sur des plans séparés

Les profils techniques sont toujours très recherchés avec

78%

d'intention de recrutement sur les Data Engineers et Machine Learning Engineers

Seules

29%

des directions data pilotent leur activité en corrélation avec des métriques métiers

44%

des sondés considèrent leur organisation data comme un frein

- Partie 1 -

Des résultats économiques encourageants, ponctués de quelques désillusions

Les investissements dans la data et l'IA semblent porter leurs fruits

Une bonne dynamique d'investissement

Les entreprises continuent d'investir dans la data malgré une hausse de certains coûts unitaires (rémunération des compétences, coût des plateformes, formation continue). Cette dynamique permet à de nombreux directeurs data de bénéficier de moyens à la hauteur de leurs ambitions.

Des budgets bien dimensionnés

77% des sondés estiment que leur budget est bien dimensionné et qu'il ne constitue pas un obstacle dans leur trajectoire.

Nous avons délivré suffisamment de valeur pour ne plus être freinés par le budget. 🗨️

Aymeric VEYRON,
Head of Actuarial, Leocare

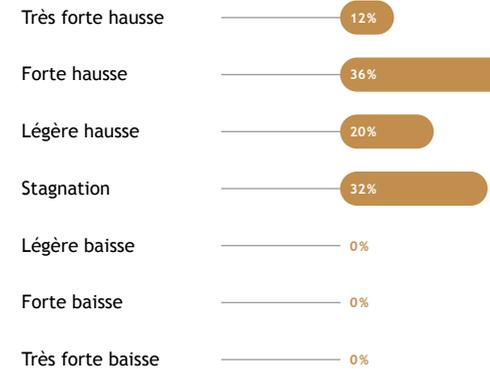
Notre Direction Générale présente la data comme un axe de développement prioritaire du Groupe depuis l'année dernière. Cela dépasse la simple rhétorique car on ressent de réels effets sur les moyens. 🗨️

Jean-Marie L'HOTELLIER,
Responsable Data Factory, Verlingue

Des gains prometteurs

Les directeurs data considèrent que leur activité devient de plus en plus rentable à mesure que les usages IA, décisionnels et data management se démultiplient.

Évolution du bénéfice moyen par cas d'usage data, déclaratif, % de répondants



68%

des sondés estiment que les bénéfices moyens tirés des cas d'usages data et IA augmentent.

+1/3

des sondés estime que l'augmentation est très marquée.

Pour les autres, on constate une stagnation car la hausse des gains s'est accompagnée :

- d'une hausse de moyens (envergure des cas d'usage, niveau d'exigence, nouvelles briques techniques...)
- et/ou d'une inflation des coûts unitaires (salaires, tarifs prestataires...)

L'augmentation globale des salaires du digital (besoin de technicité et d'expérience, inflation...) impacte les coûts de nos projets data mais est pour le moment compensée par une meilleure efficacité liée à la mise en place de plateformes et de réutilisation de code. De plus, la proximité croissante des Data Scientists avec les équipes business leur permet d'accélérer le temps d'appropriation et donc le nombre de sujets traités par une scrum team. Le ROI de nos projets data ne fait que progresser ce qui nous permet de justifier l'augmentation de nos budgets IA. 🗨️

Gilles COLAS des FRANCS,
Head of AI Transformation, Michelin

Des sources de valeurs diverses, partiellement dépendantes des secteurs d'activité

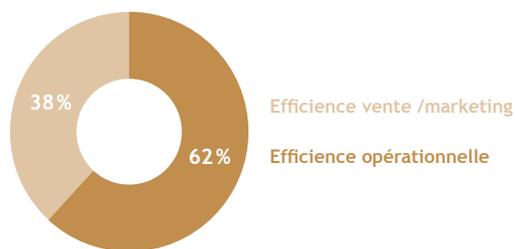
Les investissements sont rentabilisés via une variété de leviers économiques, offrant principalement des gains d'efficacité opérationnelle (62% des usages) ou marketing/vente (38% des usages).

Sans surprise, les secteurs B2C comme le CPG, le retail, le luxe, les banques et assurances tirent plus de bénéfices sur la vente/marketing. L'industrie, l'énergie et la santé trouvent davantage d'opportunités sur les opérations.

Dans l'ensemble, les portefeuilles de cas d'usage conservent un certain équilibre démontrant la capacité des équipes data à résoudre une variété de problématiques métiers, portées par différentes fonctions de l'entreprise.

Répartition des gains par levier économique

% de répondants



Exemples de leviers économiques

Efficiency vente/marketing :

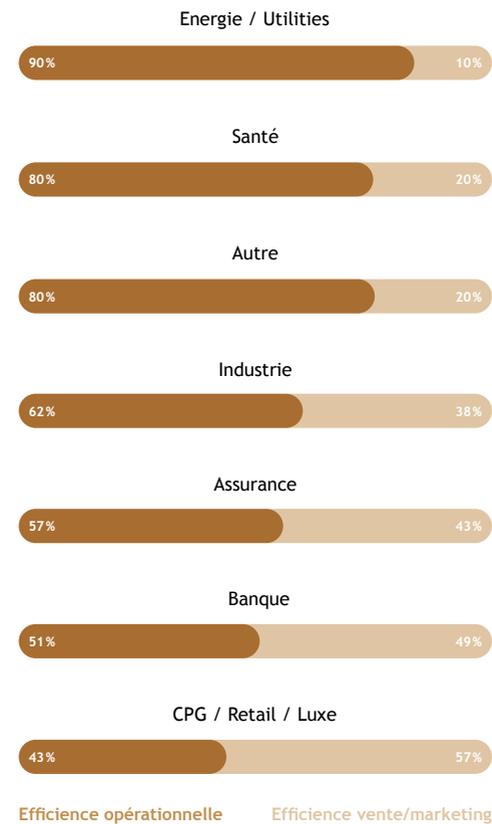
- Upsell/cross sell
- Fidélisation
- Communication/offres ciblées
- Pricing dynamique
- ...

Efficiency opérationnelle :

- Automatisation
- Optimisation de processus
- Optimisation du rendement
- ...

Répartition des usages selon le levier économique et le secteur d'activité

% moyenne



*CPG = Consumer Packaged Goods ou biens de consommation.

Beaucoup d'entreprises se recentrent sur les fondamentaux et dépriorisent notamment les projets d'IA complexes

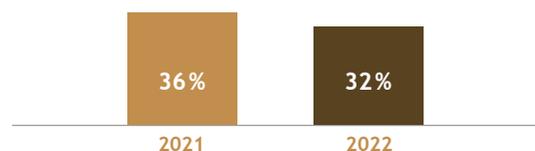
Une préférence pour les projets big data, BI et data management

En 2021-22, beaucoup de directions data ont privilégié des projets reposant sur des pratiques et des technologies maîtrisées. Par exemple :

- Création de jeux de données réutilisables : jointure, ingestion, mise en qualité et APIsation
- Mise sous contrôle de référentiels ou de données transverses
- Migration vers le Cloud
- ...

Parts de cas d'usage intégrant de l'IA

% de répondants



“ La proportion de projets qui relève de data management a augmenté. La valeur peut provenir d'un simple accès à une donnée de qualité permettant des initiatives data qui mobilisent à travers l'entreprise. ”

David ELVIRA, Head Of Digital R&D, Ipsen

“ Nous mettons la priorité sur les sujets de référentiels et sur le socle de données pour avoir des outils plus accessibles, plus ouverts et ainsi débloquer de la valeur. ”

Nicolas CORNET, Group Head of Data, Saint Gobain

Un recul de l'IA dans les projets data

La part de l'IA dans les projets data portés par les entreprises diminue de 4 points entre 2021 et 2022.

Cette baisse, d'ampleur limitée, se remarque toutefois car :

- elle n'a pas de précédent
- elle concerne la majorité des entreprises

Le marché traverse une phase de désenchantement envers l'IA, à l'issue de laquelle les entreprises gagnent en lucidité

Une phase de désillusion envers l'IA ?

Aujourd'hui, presque toutes les entreprises (93% dans notre panel) font de l'IA. Toutefois, l'engouement initial envers cette technologie donne de premiers signes d'essoufflement.

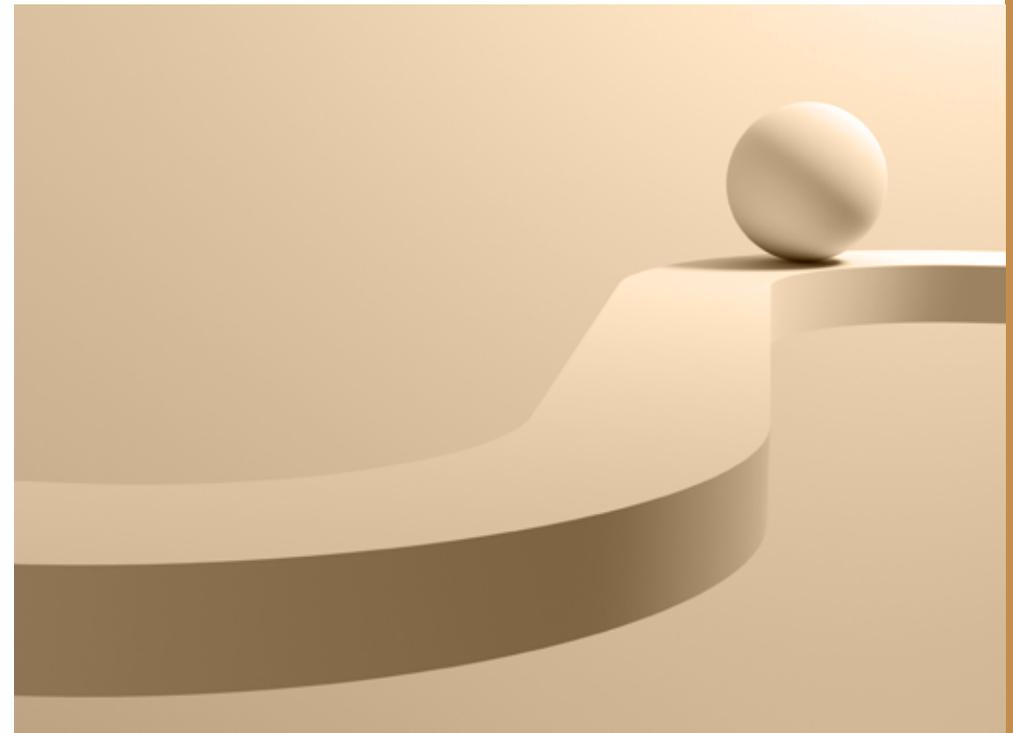
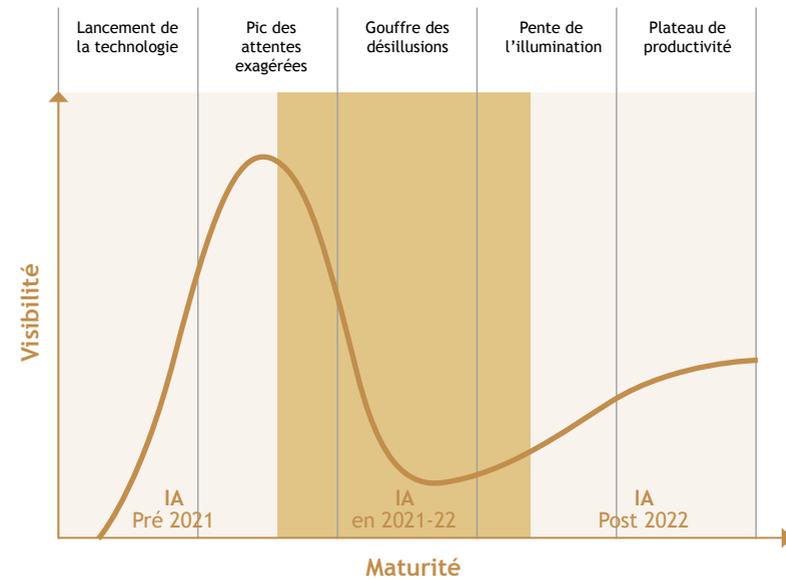
En cause : un enthousiasme qui s'est parfois accompagné de promesses démesurées et d'une course au spectaculaire.

Oui mais transitoire !

L'engouement envers l'IA a toutefois été bénéfique. Il a permis à de nombreuses entreprises de :

- réaliser les investissements à rendement différé, nécessaires pour construire les fondations, notamment des socles technologiques
- tester et développer progressivement leur capacité à faire, notamment sur l'industrialisation de l'IA
- sécuriser les talents sur un marché de l'emploi très concurrentiel
- calibrer leurs attentes en matière de création de valeur
- mieux évaluer la faisabilité des projets

Les entreprises pionnières ayant traversé la phase de désenchantement s'engagent depuis sur des projets IA offrant un meilleur ratio risque/bénéfice. En concentrant leurs efforts sur les bons cas d'usage, elles réussissent, pour la plupart, à mener des projets IA à fort impact jusqu'en phase de run.



Les entreprises qui persévèrent dans l'IA y trouvent de puissants leviers de création de valeur



SUCCESS STORY 1

Leader dans le secteur de l'Assurance

Problème

Chaque jour, cet assureur reçoit des millions d'e-mails de la part de ses sociétaires, concernant une diversité de contrats (habitation, mobilité...) et de sujets (demande d'information, déclaration de sinistre...). Ces e-mails doivent atteindre le bon interlocuteur au sein de l'entreprise pour garantir une réponse pertinente dans les meilleurs délais.

Avec un volume d'emails conséquent et en croissance continue, les délais de réponse atteignent en moyenne 3 jours. Ces délais dégradent la relation client ainsi que l'efficacité des processus internes.

Solution

Exploiter les capacités de l'IA, notamment le NLP*, pour automatiser certaines étapes du traitement des e-mails :



Routage :

orienter les e-mails vers la bonne entité



Enrichissement vue métier :

afficher un résumé de l'email pour faciliter son traitement



Priorisation des emails :

détecter des indicateurs d'urgence et d'insatisfaction pour réagir promptement



Clôture de tâches :

clôture les tâches pour les emails peu pertinents



Réponse automatique :

générer une réponse rapide aux demandes simples

Résultat

Répondre **40%** plus vite

aux sollicitations clients en routant chaque année
+ de 2,5M d'e-mails

*Traitement du langage naturel

SUCCESS STORY 2

Leader du secteur CPG/retail/ Luxe

Problème

Éprouvant des difficultés à prédire les ventes, l'entreprise oscillait en permanence entre :

- **Surproduction** : stock excessif, invendus, logistique complexe et inefficace entre les points de vente
- **Sous-production** : taux de service dégradé générant une perte de chiffre d'affaires sur les articles indisponibles en rayon

Solution

Exploiter les capacités de l'IA, notamment les séries temporelles, pour affiner les prévisions de vente :



Optimisation des réassorts magasin depuis l'entrepôt central



Modélisation du réseau logistique



Optimisation des coûts d'opération et le taux de service en magasin



Anticipation des pics de vente durant les fêtes

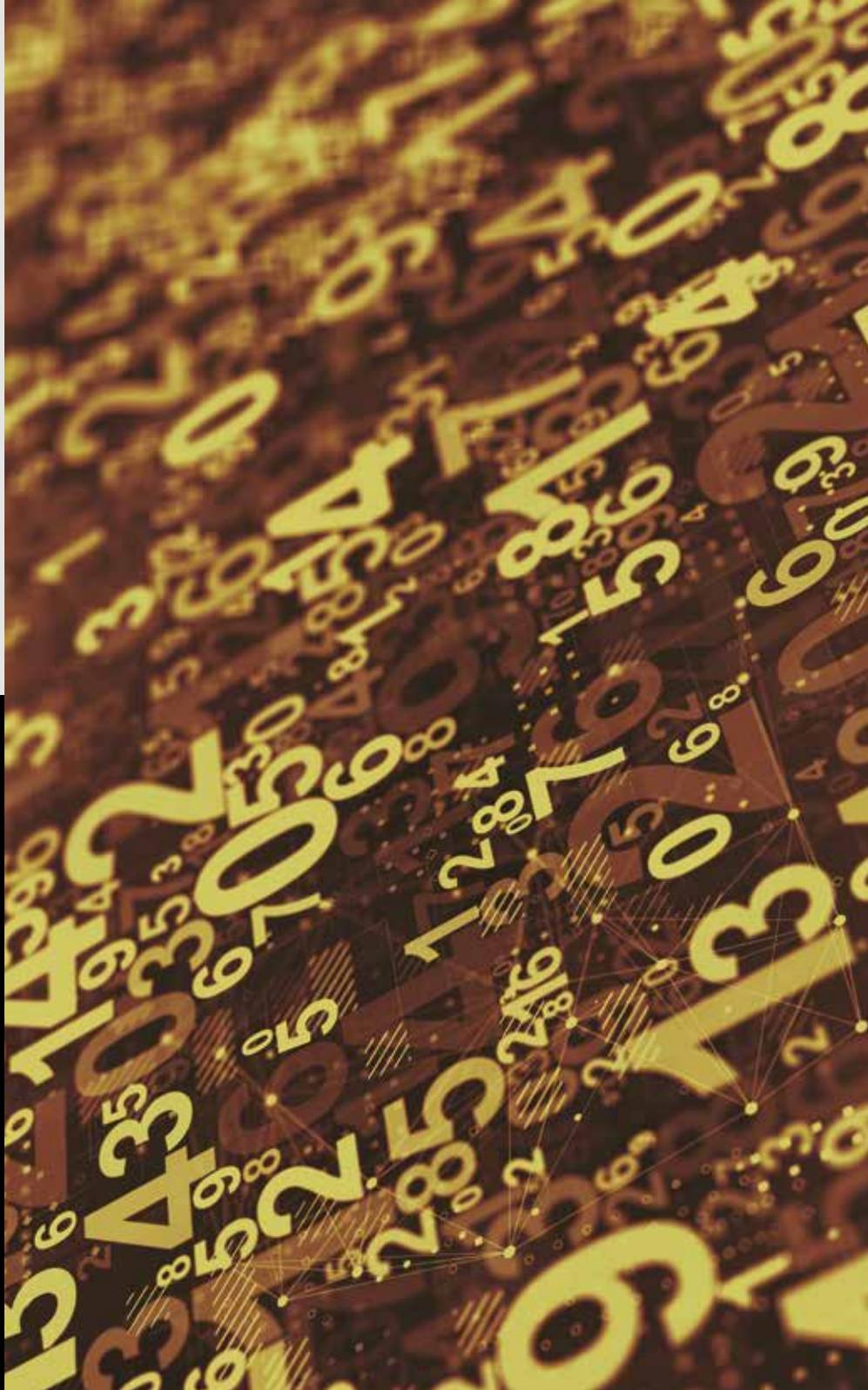


Emploi de la méthodologie MLOps pour maintenir le modèle dans le temps

Résultat

Économiser **> 80 M€** par an

en réduisant les ruptures de stocks et les opérations de réassort



Synthèse et analyse

Une bonne dynamique à maintenir en se rapprochant davantage des métiers.

Les directeurs data constatent une bonne dynamique d'investissement et considèrent que les budgets ne freinent plus leurs ambitions. Après plusieurs années de service, ils ont dans l'ensemble convaincu l'entreprise de leur valeur, et justifié les investissements réalisés par le passé.

Après avoir mené une course aux usages pour générer de la valeur rapidement, les entreprises se sont recentrées cette année sur les fondamentaux. Cette tendance s'est accompagnée d'une dépriorisation des projets IA, qui avaient obtenu une place de choix dans les feuilles de route data, lorsque les entreprises cherchaient encore à se tester.

Cette baisse ne signe pas pour autant le déclin de l'IA. Au contraire, elle s'interprète comme une étape de transition vers plus de maturité. Après les expérimentations tous azimut, les entreprises doivent à présent faire des choix plus raisonnés au regard d'un rapport coût vs bénéfice mieux évalué. Cela passera sans doute par un partenariat renforcé entre data et métiers, c'est-à-dire :

- des convictions partagées par rapport à l'IA et son potentiel
- des feuilles de route synchrones
- des équipes pluridisciplinaires à chaque étape du cycle de vie d'un projet.

- Partie 2 -

En route vers le run ! Les entreprises franchissent un nouveau cap d'indus- trialisation de l'IA

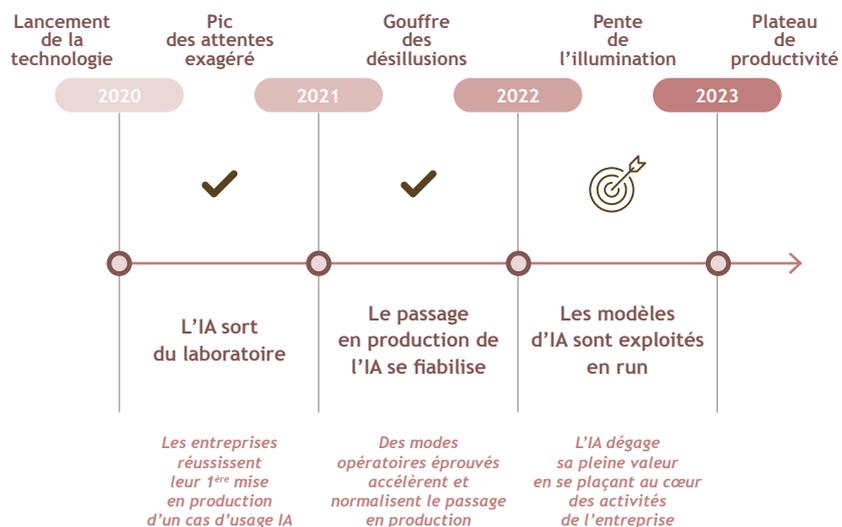
En 3 ans, les entreprises françaises ont franchi des jalons de maturité successifs

Alors qu'elles maîtrisent la Business Intelligence (reporting, dashboards...) depuis un certain temps, les entreprises conservent une marge de progrès plus importante pour les usages fondés sur l'IA (Machine Learning, Deep Learning...). Au cours des 3 dernières années, elles ont toutefois réalisé des progrès notables.

Pré 2020, l'IA se présente comme un nouveau relai d'innovation. La plupart des entreprises réalisent alors leurs premières expérimentations pour tester le potentiel de la technologie. Entre 2020 et 2021, l'IA sort du laboratoire! Les premiers cas d'usages passent en production, mais ils restent rares. En pleine période d'euphorie, les entreprises s'engagent encore souvent sur des projets voués à l'échec car trop exigeants ou pas assez générateurs de valeur.

S'en suit une phase de désillusion marquée par un recours à l'IA moins fréquent mais mieux ciblé. Les entreprises s'éparpillent moins, parient sur les projets véritablement porteurs et anticipent mieux les difficultés et l'effort à fournir.

Cette phase de désillusion qui a marqué l'année 2021-22 s'accompagne donc d'une hausse très marquée du taux d'industrialisation des projets d'IA. Disposant de solutions d'IA industrielles, les entreprises doivent à présent atteindre un nouveau jalon : débloquer la pleine valeur de leurs usages en phase d'exploitation !



Industrialisation et exploitation de l'IA : de quoi parle-t-on ?

Le tableau ci-contre clarifie les étapes de développement et d'exploitation d'un cas d'usage IA. Dans ce chapitre, nous portons une attention particulière aux étapes de d'industrialisation et d'exploitation/run.

	Temps →			
Étape	Build (peut couvrir tout ou une partie)		seulement des processus ci-dessous	
			Mise en production	Run
Sous-Étape	Expérimentation		Industrialisation	
Processus	Screening Cadrage - POC	Prototypage	Développement pilote	Exploitation
Objectif	Évaluer la faisabilité, la pertinence du modèle et la difficulté du développement	Faire valider l'apport de valeur par le métier	Démontrer une valeur métier en contexte d'utilisation réel	Pérenniser et améliorer le ROI (maintenance, nouvelles features...)
Investissement	Négligeable	Moyen	Élevé	Très élevé
Environnement	Environnement local	Environnement de Dev	Environnement de recette	Environnement de production
Données input	Échantillon restreint de données	Simuler l'usage de l'IA sur des données froides d'échantillon	Flux manuel ou partiellement automatisé	Flux entièrement automatisé
Utilisateurs	Client/sponsor	Testeur	Key users	End users
UX (optionnel)	Maquette	Interface rudimentaire	Interface v0	Interface évolutive intégrée aux systèmes opérationnels de prise de décision

Entre 2020 et 2021, l'IA sort du laboratoire

Entre 2020 et 2021, près de neuf entreprises sur dix ont mis en production au moins un cas d'usage IA

3 accélérateurs principaux ont permis cet accomplissement :

- Des stratégies volontaristes sur l'IA soutenues par un Senior Management mieux acculturé et désireux d'explorer des opportunités
- La disponibilité de plateformes data facilitant l'ingestion, le traitement et l'exposition de données en dehors des silos applicatifs et organisationnels
- Le recours croissant à des profils hybrides data/IT, comme le Data Engineer, pour fluidifier le passage de relais entre DSI et Data Lab

Nous avons besoin d'une bonne infrastructure pour passer le cap de la mise en production. Notre Data Lake porte aujourd'hui plusieurs cas d'usages industrialisés. 🗨️

Bertrand NACHBAUR,
Chief Information Officer, Dalkia

On a formé des Data Scientists au Data Engineering pour qu'ils prennent en compte l'industrialisation dès le début d'un projet. 🗨️

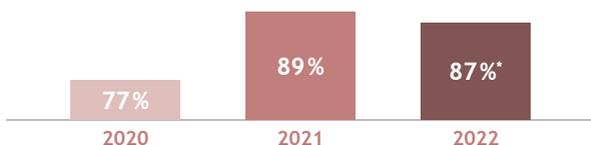
Sylvain CHAMPENOIS,
Data Team Leader, Natixis

La data est un sujet qui remonte jusqu'au COMEX. Elle fait partie des 15 axes de réflexion stratégique qui orientent notre trajectoire à travers le Groupe. 🗨️

Olivier URCEL,
Chief Data Officer, Stellantis

Part d'entreprises ayant passé au moins un cas d'usage IA en production

% de répondants



*81% si on inclut les data offices ayant moins d'un an d'existence

Entre 2021 et 2022, le passage en production de l'IA se fiabilise

Une amélioration généralisée du niveau d'industrialisation au cours de l'année 2021-22

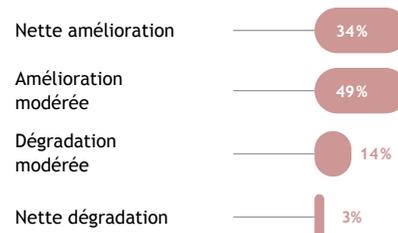
- **83%** des entreprises déclarent avoir augmenté leur capacité d'industrialisation
- Le taux d'industrialisation moyen des cas d'usage IA bondit de 27% à **46%**

L'atteinte d'un seuil optimal pour de nombreuses entreprises

- **39%** des sondés passent plus d'un cas d'usage IA sur deux en production
- Dont **7%** des sondés qui les passent tous sans exception

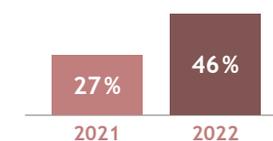
Evolution de la capacité à industrialiser un projet IA

Déclaratif, % de répondants



Evolution du taux d'industrialisation des projets IA

% moyen



À présent, on ne fonctionne plus en mode POC.

L'industrialisation est prise en compte dès la phase de cadrage. 🗨️

Karim LOUEDEC,
Group Chief Data Officer, Servier

Répartition du panel selon le taux d'industrialisation des projets IA

% de répondants



Une maîtrise qui se développe progressivement

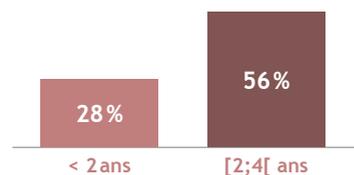
Il faut, en moyenne, deux à quatre années d'investissements et d'expérimentation pour qu'une entreprise franchisse la barre des 50% d'usages industrialisés.

Une marche à franchir qui se réduit pour les nouveaux entrants

Les directions data les plus récentes (moins de deux ans) semblent bénéficier de l'expérience accumulée par les pionniers. Leur taux moyen d'industrialisation s'élève à 28%, ce qui correspond au niveau observé sur l'ensemble du marché en 2021.

Taux d'industrialisation des projets IA selon l'âge de l'entité data

% moyen



Pour nous, un modèle d'IA est industrialisé si (a) il est APIisé et mis à disposition dans notre SI et (b) s'il fait partie intégrante des processus métier. 95% de nos cas d'usage IA remplissent ces critères. 📍📍

Romain MERIDOUX, Head of Data Lab, CNP Assurances

Nous n'avons plus de difficulté à industrialiser. 📍📍

Bruno GANEVAL,
Chief Data Officer, RCI

Notre cible est de passer au minimum 50% de nos usages en production. 📍📍

Hind MECHBAL,
Chief Information Officer, CCR

De nombreuses entreprises accélèrent également leur industrialisation grâce aux outils Low-Code/No-Code ainsi qu'aux solutions logicielles

Low Code/No Code (ex : Dataiku, Alteryx, data robot)

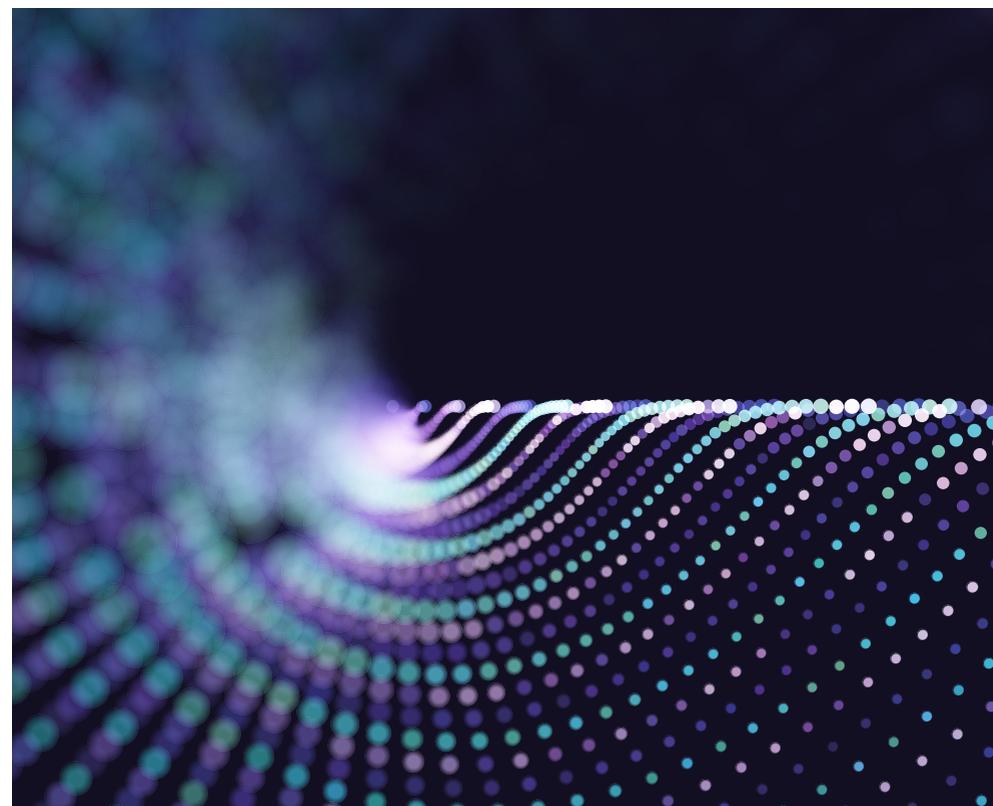
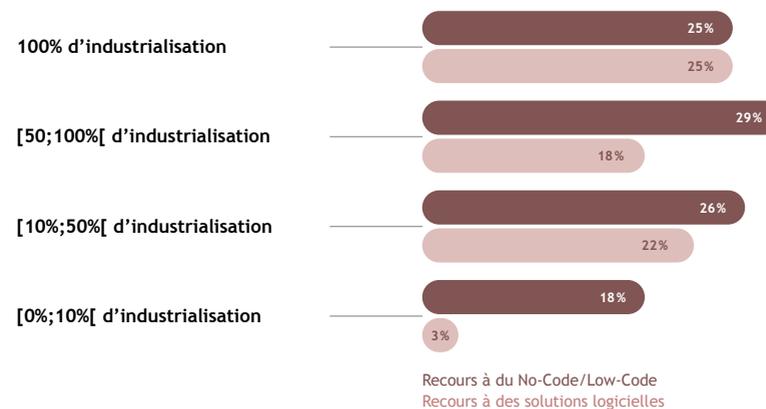
- En moyenne, un quart des cas d'usage s'appuie sur ces outils.
- En plus de démocratiser l'analyse des données, ils permettent souvent de fluidifier le passage des cas d'usage en environnement de production.
- Les cas d'usage s'avèrent toutefois limitants pour certains usages à plus fort enjeu scientifique. Certaines entreprises commencent à formaliser leurs politiques de recours à ces plateformes.

Solutions logicielles (ex : chatbots, micro-services IA...)

- 12% des cas d'usage se fondent entièrement ou en partie sur des solutions logicielles achetées sur le marché.
- Le recours aux solutions du marché augmente parallèlement au taux d'industrialisation des cas d'usage.
- Elles sont toutefois moins plébiscitées par les entreprises de faible maturité (3% des cas d'usage en dépendent). Celles-ci ont un choix de solutions très limité sur le marché, dès lors qu'elles ne sont pas capables d'industrialiser leurs flux de données et d'en assurer la qualité.
- Globalement, les arbitrages de make or buy se formalisent.
Plus de 50% des entreprises suivent une stratégie explicite à ce sujet.

Comparaison des taux d'industrialisation des cas d'usage intégrant de l'IA selon le recours aux solutions sur étagère et aux outils de No-Code ou Low-Code

% de répondants



Le prochain défi : exploiter pleinement les cas d'usage IA en run

De meilleures capacités techniques pour maîtriser l'IA en run

Entre 2021 et 2022, la part des entreprises pratiquant le MLOps est passée de **47% à 72%**. Il existe toutefois des différences de maturité importantes parmi les pratiquants :

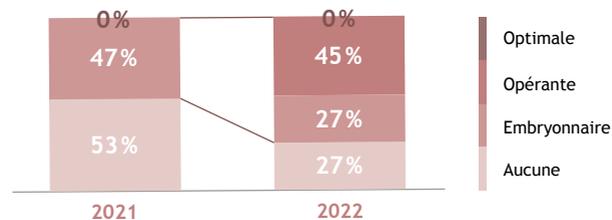
- **38%** des pratiquants se considèrent encore à un stade embryonnaire
- **62%** trouvent leur pratique opérante mais pas encore optimale

Des freins organisationnels qui persistent

En 2020, **70%** de sondés* avaient initié des réflexions sur les questions organisationnelles et budgétaires relatives à l'exploitation des modèles IA. Deux ans plus tard, beaucoup d'entreprises peinent encore à délimiter les responsabilités des directions data, de la DSI et des métiers concernant le maintien, le contrôle et le financement des usages en run.

Evolution de la maturité déclarée des entreprises pour la pratique du MLOps

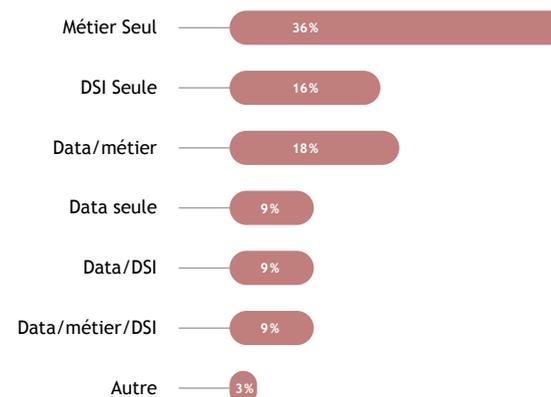
% de répondants



*Baromètre des Directions Data 2021

Mode de financement des usages data en run

% de répondants



Sur la question du financement, les entreprises avancent progressivement vers une forme de consensus. Le métier doit porter les coûts et les ressources, que se soit seul (36% des sondés) ou de façon partagée avec la DSI et/ou le data office (36%).



On atteint 90% d'industrialisation sur le SI, mais si l'on parle de rôles et responsabilités en phase de run, cela devient plus compliqué. Le Data Office ne peut pas porter tous les modèles en run. Ce n'est pas scalable. ●●

Maxime HAVÉZ,
Chief Data Officer, Arkea

Un prérequis pour exploiter sereinement l'IA en run : l'IA de confiance !

A mesure que l'IA se propage dans les processus métiers, le besoin de confiance et de maîtrise s'accroît au sein des entreprises. En phase de run, les travaux autour du cycle de vie des modèles devront prendre en compte les enjeux de l'intelligence artificielle de confiance.

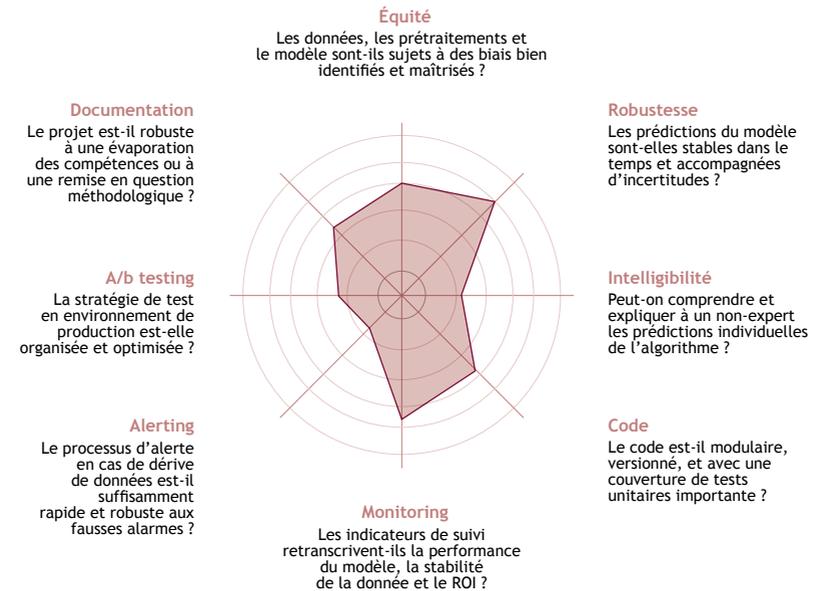
Si plus de **80% des entreprises** établissent déjà des contrôles sur leur modèles de Machine Learning, peu couvrent l'ensemble des axes d'une IA de confiance. Les enjeux perçus comme prioritaires pour la mise en place de ces contrôles sont :

- 1 - de protéger l'image de l'entreprise d'éventuels biais ou erreurs (ex : discrimination)
- 2 - d'assurer la robustesse, la fiabilité et la précision des prévisions
- 3 - rendre les modèles intelligibles et transparents

De plus, la stratégie et l'organisation autour des enjeux de confiance restent encore à construire pour la plupart des entreprises. En 2021, elles étaient seulement **34%** à avoir établi une charte explicite sur ces sujets.



Exemples de critères d'évaluation d'un modèle IA



Auditer les modèles d'IA

Des contrôles internes doivent être réalisés sur les IA en run de façon régulière, selon un plan de contrôle prédéfini et adapté à chaque usage. Ce plan doit être cadré en amont du projet et mis en œuvre sur l'entièreté du cycle de vie du modèle.

En plus de renforcer la confiance des parties prenantes, mener des audits réguliers permet de cadencer les opérations de maintenance et d'amélioration continue. Leur récurrence assure que le modèle suive les évolutions ayant lieu sur les processus, la réglementation ou les exigences internes.

Pour être exhaustive, la démarche d'amélioration continue doit porter un volet organisationnel mais aussi technologique. En effet, le domaine de l'IA de confiance évolue rapidement et voit notamment l'apparition régulière de nouvelles méthodes d'intelligibilité, de maîtrise des biais ou de sobriété.

IA de confiance : de quoi parle-t-on ?

Le 20 octobre 2020, le parlement européen a adopté à une grande majorité un texte sur la réglementation de l'IA. Ce texte confirme que la réglementation sur l'IA de confiance sera fondée sur 7 exigences clés et que celle-ci s'appliquera uniquement aux IA dites « critiques ».

En effet, le souci de soutenir les entreprises dans la mise en place de leur IA de confiance en évitant de brider l'innovation, conduit la Commission Européenne à légiférer non pas sur les technologies elles-mêmes mais sur les domaines d'application où l'IA est critique. Pour les IA non critiques, la Commission se limite à inciter les entreprises et organisations à suivre ses recommandations.

Secteurs à haut risque

Secteur public*

Défense et sécurité

Finance, banque et assurance

Emploi

Éducation

Soins de santé

Transports

Énergie

*Asile, migration, contrôles aux frontières, système judiciaire et services de sécurité sociale

Usages ou finalités à haut risque

Recrutement

Notation et évaluation des étudiants

Affectation de fonds publics

Octroi de prêts

Commerce, courtage, fiscalité, etc.

Traitements et procédures médicaux

Processus électoraux et campagnes politiques

Gestion des déchets

Décisions du secteur public ayant une incidence importante et directe sur les droits et obligations des personnes physiques ou morales

Conduite automatisée

Gestion du trafic

Systèmes militaires autonomes

Production et distribution d'énergie

Contrôle des émissions

Synthèse et analyse

Pour réussir à l'ère du run, les entreprises devront relever de nouveaux défis

Les entreprises françaises ont franchi un véritable cap dans leur transformation data. En passant d'une approche expérimentale à une approche industrielle de l'IA, elles ont concrétisé les promesses de cette technologie.

Cela dit, la transformation doit se poursuivre pour que la valeur soit pleinement réalisée. Après l'industrialisation, un modèle doit être maintenu au bon niveau de service; sa performance et son utilisation doivent être surveillées; et les retours d'utilisateurs doivent être pris en compte dans une démarche d'amélioration continue. Or aujourd'hui la bonne adoption des projets data par les métiers n'est que très rarement mesurée et les équipes data ne sont ni organisées ni outillées pour suivre l'ensemble du cycle des projets.

Les entreprises qui réussissent leur phase d'exploitation ont souvent changé de paradigme pour adopter une logique de « produit data » et non de projet ayant par définition une date de fin. On constate notamment 3 bonnes pratiques :

– 1 –

Faire intervenir les utilisateurs au plus tôt dans la conception des cas d'usage, pour s'assurer que les métiers s'approprient les modèles et qu'ils les inscrivent réellement dans leurs processus

– 2 –

Prendre en compte le cycle de vie dès la phase de cadrage, pour anticiper au mieux les enjeux techniques, scientifiques et métiers qui se poseront après la mise en production

– 3 –

Disposer d'équipes dédiées pour assurer un suivi continu de chaque cas d'usage, avec une répartition claire entre DSI, data et métiers.

- Partie 3 -

Trois défis à relever pour générer plus de valeur

Trois défis à relever pour générer plus de valeur

Pour continuer de dégager de la valeur avec la data et exploiter pleinement l'IA en phase de run, les entreprises ont placé 3 défis au centre de leurs priorités.

Défi #1

Faire aboutir le modèle organisationnel

Défi #2

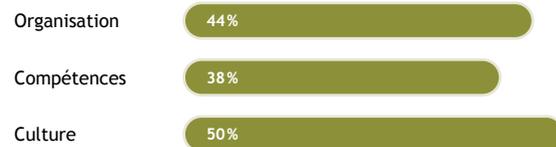
Sécuriser des compétences sous tension

Défi #3

Comblent le fossé culturel entre data et métiers

Top 3 des facteurs cités comme freins à l'industrialisation

% de répondants

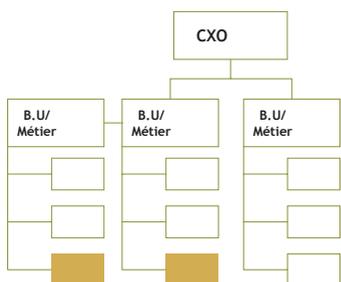


Défi #1

Faire aboutir le modèle organisationnel

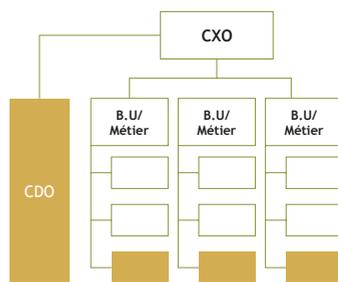
Les organisations data se répartissent en 3 modèles, chacun présentant ses propres avantages et inconvénients.

Organisation en silo



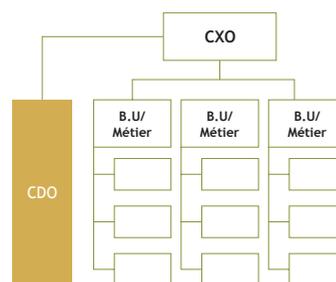
- Ressources data
- + Proximité avec le métier
- + Réactivité pour la réalisation des cas d'usage
- Faible scalabilité du modèle
- Peu de collaborations transverses

Modèle hybride



- Ressources data
- + Scalabilité du modèle
- + Standardisation déclinée dans chaque métier
- + Capitalisation et mutualisation des ressources
- Complexité du modèle et besoin récurrent d'alignement

Opérateur centralisé



- Ressources data
- + Scalabilité de processus standards
- + Capitalisation des expériences
- + Mutualisation des moyens
- Eloignement des besoins métiers
- Manque d'agilité pour la réalisation des cas d'usage

84%

des entreprises ont établi ou s'orientent vers des modèles d'organisation hybride

Défi #1

Faire aboutir le modèle organisationnel

À mesure qu'elles mûrissent, les entreprises se tournent quasi unanimement vers les modèles hybrides

Les modèles en silo

sont rares et constituent généralement une étape transitoire pour les entreprises qui débutent leur transformation data. Souvent c'est un métier pionnier, déjà producteur de cas d'usage data, qui concentre mécaniquement les ressources humaines et technologiques permettant leur réalisation.

Les modèles centralisés

sont quant à eux privilégiés par les entreprises qui commencent à étendre leur transformation dans l'entreprise. La centralisation permet aux métiers de débiter sans mobiliser eux-mêmes des investissements souvent dissuasifs. Avec le temps, ces modèles atteignent leurs limites : goulots d'étranglement, éloignement des sujets à forts enjeux.

Les modèles hybrides

sont alors privilégiés pour offrir aux métiers une capacité à piloter leur propre feuille de route data, tout en maintenant une cohérence avec le reste de l'entreprise. L'hybridation demande toutefois un leadership fort de la part de l'entité data centralisée. Elle nécessite aussi, pour beaucoup d'entreprises, de faire évoluer les architectures héritées de modèles centralisés.

Modèles d'organisation adoptés par les entreprises selon l'âge de la direction data

% de répondants



Modèle d'organisation

■ Silo ■ Hybride ■ Centralisé

L'adoption de modèles hybrides facilite l'industrialisation

Les entreprises qui adoptent un modèle hybride sont aussi celles qui réussissent le mieux l'industrialisation de l'IA. Près d'un cas d'usage IA sur deux passe en production avec ce modèle d'organisation, contre 34% pour les modèles centralisés et 28% pour les modèles en silo.

Le score de satisfaction sur la fluidité du processus de delivery est également le plus élevé, avec une note moyenne de 7/10 contre 6/10 pour les autres modèles.

Les avantages du modèle hybride les plus souvent cités sont :

- Capitalisation sur un patrimoine commun (technologie, données, compétences) développé dans un écosystème data étendu
- Identification et diffusion des bonnes pratiques fluidifiant le delivery
- Capacité à investir dans des plateformes data performantes au service d'une base d'utilisateurs étendue
- Mise à disposition de ressources gérées par le data office en soutien aux métiers moins matures

Capacité d'industrialisation atteinte selon le modèle d'organisation data adopté

Part des usages IA mis en production, % en moyenne

Processus de Delivery (score de satisfaction de 0 à 10)



L'implémentation des modèles hybrides demeure toutefois inaboutie pour beaucoup d'entreprises

L'organisation data hybride est vertueuse, mais en pratique, elle reste complexe et laborieuse à mettre en oeuvre. Les directeurs data ainsi que leurs homologues métiers et DSI doivent s'accorder sur des questions particulièrement structurantes.

Coté Data Gouvernance :

- Quelle répartition des rôles, des moyens et des incitations entre le central et les relais data ?
- Quelle légitimité pour le central à établir des règles, des normes, ou des recommandations ?
- Comment casser les silos préexistants dans l'organisation ?
- Comment harmoniser les pratiques d'un réseau data souvent hétérogène ?
- Comment coordonner les feuilles de route de la communauté data sans entraver l'autonomie de ses membres ?

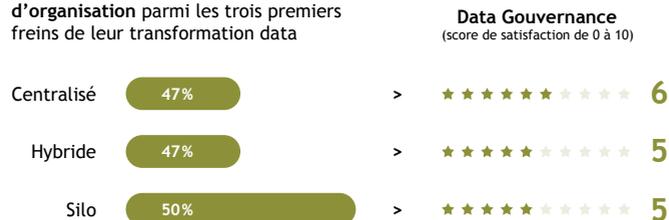
Coté Architecture & Plateformes :

- Comment permettre une meilleure circulation de la donnée alors que l'architecture data existante est fragmentée ?
- Comment assurer l'interopérabilité des modèles développés par différentes entités ?
- Comment permettre aux métiers de produire et consommer des données de qualité sans passer par une équipe de data engineering centralisée ?

En l'absence d'accord et de feuille de route concrète pour répondre à ces questions, près de la moitié des entreprises se déclarent ralenties dans leur transformation data.

Insatisfaction des entreprises pour leur modèle d'organisation data

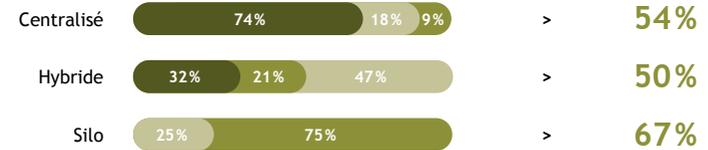
% d'entreprises citant des problèmes d'organisation parmi les trois premiers freins de leur transformation data



Comparaison des architectures data selon le modèle d'organisation

% de répondants

Répartition du panel selon l'architecture et le modèle d'organisation data



Part des répondants qui souhaitent faire évoluer leur architecture prochainement

Architecture Fragmentée : multiplicité de briques de stockage, de calcul et de développement, chacun servant des données ou un usage particulier

Architecture Hybride : briques de stockage, de calcul et de développement déployées individuellement par chaque métier suivant des patterns communs imposés

Architecture Centralisée : briques de stockage, de calcul et de développement, communes à plusieurs entités et gérés exclusivement par une équipe dédiée

Pour que leur modèle opérationnel fonctionne à l'optimum, les entreprises doivent aligner leur modèle d'organisation data à leur architecture. Cependant, à date, plus de la moitié des organisations hybrides reposent sur un socle inadapté : 32% sur une architecture hyper centralisée (ex : data lake géré par une entité centrale) et 21% sur une architecture fragmentée (ex : multiples data warehouses non communicantes). Les 47% restantes sont déjà en train d'aligner leur architecture sur un principe de modèle hybride.

Pour répondre à cet enjeu de mise en cohérence de la gouvernance et de l'architecture avec une organisation hybride, certaines entreprises se tournent vers les principes du Data Mesh. Si aujourd'hui seulement 13% déclarent l'adopter, le Data Mesh apparaît comme un mouvement de fond qui devrait continuer à se développer ces prochaines années.

Le Data Mesh n'est pas qu'une mode, il faudra qu'on y aille pour rendre les métiers plus autonomes sur l'usage de leurs données. ”

Chief Data Officer

Des plateformes technologiques à faire évoluer



Nous avons pour objectif de décloisonner l'usage des données à travers la banque. Nous étudions à ce titre des capacités de data virtualization sur nos Systèmes d'information ainsi que la mise en place d'un datahub réconciliant les mondes opérationnels et décisionnels, afin de permettre la mise en place d'IA à l'échelle. ””

Matthieu OLIVIER,
Chief Data Officer, La Banque Postale



On s'oriente de plus en plus vers du Data Mesh, même si le terme n'est pas encore entièrement officialisé dans toutes les instances. ””

Chief Data Officer



Historiquement, notre Datalake fonctionnait de façon hypercentralisée. C'est le Groupe qui poussait les données aux filiales. Nous avons évolué vers plus de décentralisation pour résoudre certaines problématiques d'accessibilité à la donnée. A présent, ce sont les filiales qui vont chercher par elles-mêmes les données dont elles ont besoin en s'appuyant sur Dataiku. ””

Chief Data Officer



Nous allons vers le Data Mesh sur les domaines référentiels, qui sont consommés par de multiples entités à travers le Groupe. ””

Chief Data Officer

Data Mesh, de quoi parle-t-on ?

La notion de Data Mesh circule depuis plusieurs années dans la presse et dans les comités de direction data et SI. Elle divise à présent le marché en trois catégories : les convaincus qui implémentent le modèle, les curieux qui attendent les premiers succès pour se décider, et les sceptiques qui ont déjà enterré le sujet.

Quel que soit son positionnement sur la question, il est important d'éviter un écueil encore commun sur le marché : traiter le Data Mesh comme une innovation technologique.

Le Data Mesh est avant tout une transformation organisationnelle visant à décentraliser le data management (self-service, autogestion des données par le métier...) tout en offrant la cohérence nécessaire pour que les données circulent sans entrave à travers l'organisation (standards, socles communs, interopérabilité...).

D'un point de vue technologique, le Data Mesh ne repose pas sur des composants d'architecture particulièrement innovants (ex: Data catalogues, ETLs, Datalakes...). C'est l'exploitation de ces technologies au service des 4 principes fondamentaux décrits en page suivante, qui représente la réelle innovation.

Puisque le Data Mesh est avant tout un enjeu d'organisation, il doit mobiliser autant les métiers et la direction data que la DSI; depuis le découpage de l'entreprise en data domains jusqu'à la création des patterns d'architecture.

4 principes fondamentaux*

1/ Domain Ownership

- Déléguer la gestion des données aux métiers les plus proches du contexte d'utilisation
- Promouvoir les échanges de pair à pair sans intermédiation

2/ Data as a Product

- Responsabiliser des équipes pluridisciplinaires sur le cycle de vie d'une donnée
- Promouvoir une utilisation fiable et autonome notamment via des APIs, et des SLAs

3/ Federated Computational Governance

- Equilibrer l'autonomie des Data Domains et l'imposition de normes collectives (interopérabilité, sécurité...)
- Favoriser l'implémentation des normes directement dans « le code » des Data Products

4/ Self Serve Data Infrastructure

- Permettre aux utilisateurs de construire leur stack technique sur un socle commun garantissant l'interopérabilité
- Offrir des capacités centralisées utiles à tous (ex : catalogue de données)

* Source : Zhamak Dehghani, Analyse Quantmetry

Défi #2

Sécuriser des compétences sous tension

Les intentions de recrutement restent élevées, particulièrement pour les profils techniques

38% des sondés estiment que les compétences dont ils disposent sont insuffisantes ou mal adaptées à leurs besoins.

Profils de delivery que les entreprises cherchent majoritairement à pourvoir :

- Le profil de Data Engineer reste sous tension, même si les intentions de recrutement reculent par rapport à 2021. 53% des entreprises recrutent pour ce poste contre 66% l'an passé. Cette apparente baisse est en réalité un basculement des besoins vers des profils hybrides de Machine Learning Engineers. Alors que ce poste était marginal l'an passé, 25% des sondés souhaitent à présent en recruter pour maîtriser le run de leurs modèles d'IA.
- Les Data Scientists sont toujours recherchés mais les projets sur lesquels ils sont positionnés ne sont plus exclusivement de l'IA/Machine Learning. Ce rôle se rapproche de plus en plus de celui du Data Analyst.
- Le besoin en compétences de pilotage stratégique et en data gouvernance semble se combler progressivement. 23% des sondés souhaitent recruter ce type de profils contre 52% en 2021. Cette diminution provient aussi d'un recours accru au recrutement interne via l'up-skilling, pour ces profils nécessitant souvent une connaissance préalable de l'entreprise.

Evolution des intentions de recrutement data par poste % de répondants

	2022	2021	2020
Data Engineering	53%	66%	39%
Data Science	40%	33%	23%
Data Analyse	28%	n.a	n.a
Architecture Data	21%	36%	9%
Stratégie Data/Gouvernance	23%	52%	9%
Gestion de projet	26%	28%	26%
Machine Learning Engineering	25%	n.a	n.a
Autre	13%	28%	14%



La rareté des Machine Learning Engineers en font une cible prioritaire dans notre appel aux ressources externes. ●●

Chief Data Officer

Défi #2

Sécuriser des compétences sous tension

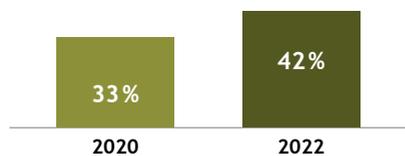
Les entreprises continuent à s'appuyer sur les ressources externes pour sécuriser les profils rares et/ou absorber une charge de travail croissante

Le recours aux ressources externes va probablement continuer de s'accroître

- Les entreprises complètent en moyenne 42% de leurs effectifs par des prestataires externes contre 33% en 2021
- Les prestataires sont principalement un moyen d'avoir accès à des expertises rares temporairement et/ou de combler des besoins capacitaires sur des profils déjà existants
- La plupart des sondés (66%) font appel à des cabinets de conseil
- Le recours aux freelances émerge en parallèle (25% des sondés)

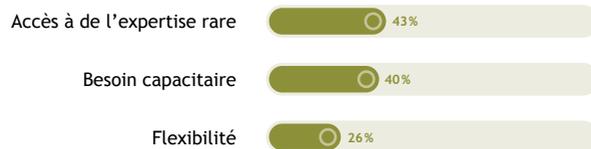
Evolution de la part d'externes dans les effectifs data totaux

% moyen



Top 3 des motifs de recours aux ressources externes

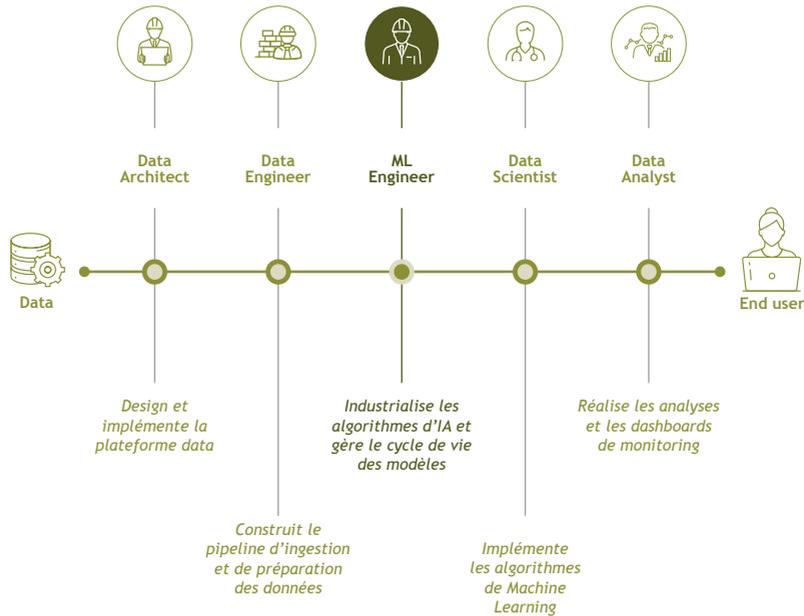
% des sondés



Machine Learning Engineer, de quoi parle-t-on ?

Responsable de la mise en production des modèles de Machine Learning/Deep Learning, le MLE est capable de connecter les équipes Data Science avec les équipes IT, de concevoir des pipelines de données fiables, de créer des interfaces aux modèles et de suivre les analyses de performances des modèles en production pour gérer les réentraînements.

Le MLE au cœur du processus d'industrialisation



Comparaison des rôles au sein d'un projet IA

	 Data Scientist	 Machine Learning Engineer	 Data Engineer
Rôle	Construire des modèles qui aident les entreprises à obtenir de meilleurs insights et à faire des prévisions à partir de leurs données	Automatiser les processus de ML et faire fonctionner les modèles dans un environnement de production	Construire, tester et maintenir les pipelines de données; fournir aux modèles ML des données de qualité
Compétences mobilisées dans le projet IA	Mathématiques et statistiques Transposition d'un problème métier en modèle data Algorithmie, prise de décision et optimisation des données Requêtage SQL Script (R/Python)	Maîtrise le cycle de vie d'un modèle et les bonnes pratiques de delivery Suivi des data et model drifts Application des méthodes Ops (CI/CD) Techniques de validation, exposition et monitoring des modèles	Script sur commandes Linux Modélisation de données Construction et monitoring de pipelines Data quality Maîtrise des technologies du cloud computing Compétence élevée en SQL
Outils principaux	Python, R, Pandas, Jupyter notebooks, SQL	Python, PyTorch, TensorFlow, MLFlow, Kubeflow, Sagemaker	SQL, Oracle, Hadoop, Amazon S3, Python

Un profil polyvalent offrant de multiples bénéfices

Le Machine Learning Engineer cumule une partie des compétences d'un Data Scientist et d'un Data Engineer. Il peut par conséquent intervenir sur l'intégralité du cycle de développement d'un modèle.

MLE - de quoi parle-t-on ? (suite)

Cette polyvalence permet de constituer des équipes de delivery plus souples et mieux armées pour l'industrialisation

Dans une équipe traditionnelle, la phase de build nécessite une implication accrue d'un ou plusieurs Data Scientist(s) : choix des algorithmes, features engineering, entraînement, comparaison des performances...

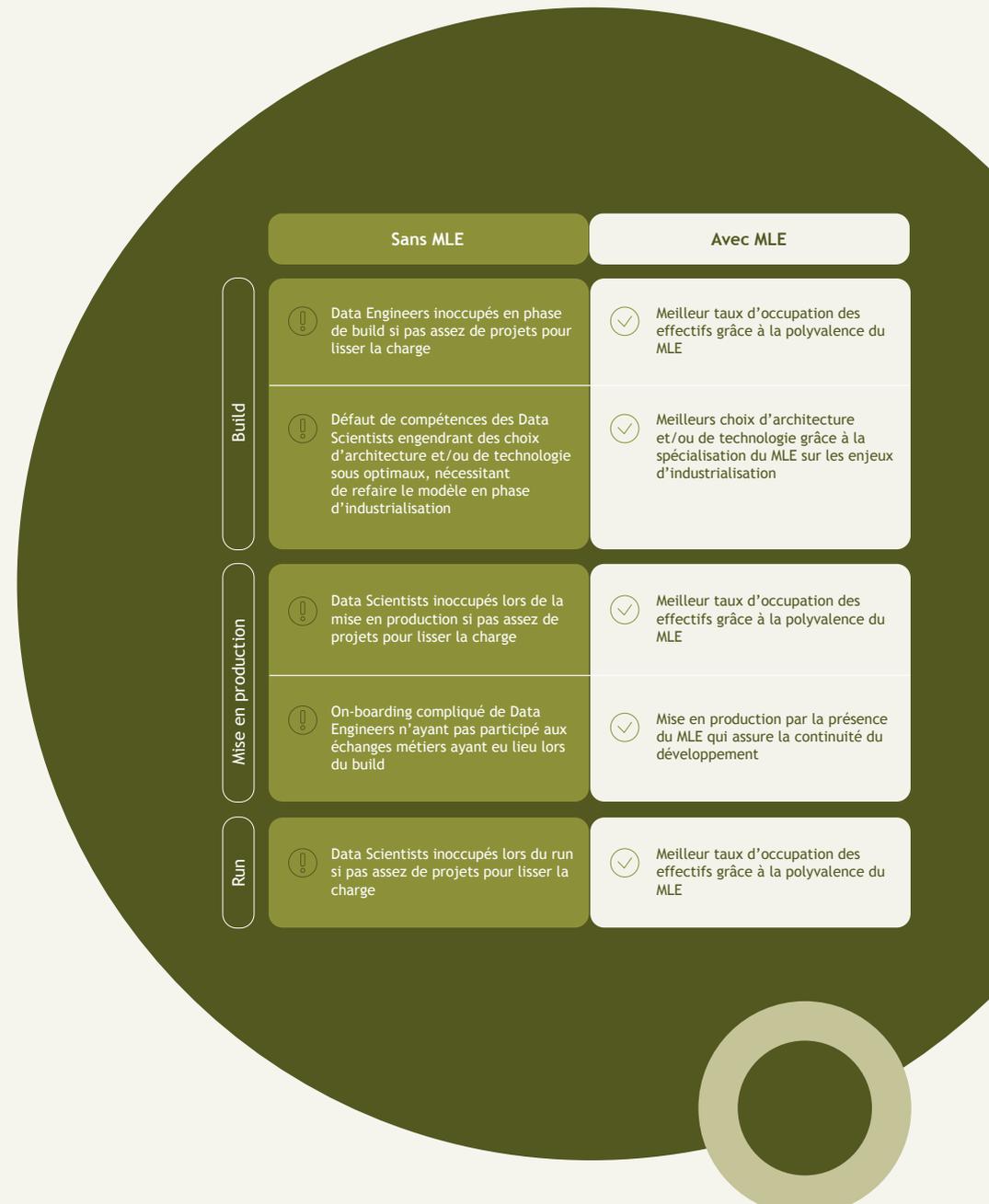
Lors du passage en production, ce sont les Data Engineers qui entrent en scène pour définir le SLO (service level objectives)/SLA (service level agreement)/SLI (service level indicators) et mettre en place les pipelines de Continuous Integration (CI)/Continuous Delivery (CD)/Continuous Testing (CT). Le modèle en run est ensuite maintenu par les opérations avec l'intervention ponctuelle de Data Engineers (monitoring et gestion des pipelines, bases de données, ETLs, APIs...) et de Data Scientists (réentraînement, ajout de features). Cette façon de travailler présente certains désavantages décrits dans le tableau ci-contre.

Une équipe intégrant un Machine Learning Engineer évite ces écueils en tirant profit de sa polyvalence. Au lieu de basculer, souvent trop abruptement, d'une phase build - centrée sur les Data Scientists - à une mise en production - centrée sur les Data Engineers - l'équipe repose sur un Machine Learning Engineer capable d'assurer la continuité du développement et des attentes utilisateurs.

Par ailleurs, la présence du Machine Learning Engineer permet d'anticiper les enjeux de l'industrialisation dès la phase de build; et de faire des choix d'architecture et de technologie pertinentes dès le début du projet.

Cela permet aux modèles de passer en production, sans être retravaillés voire repris de zéro, comme c'est souvent le cas sans MLE.

Enfin, le MLE peut intervenir en binôme soit avec un Data Engineer (si l'infrastructure est complexe par exemple) soit avec un Data Scientist (si le modèle est complexe et l'infrastructure simple), ce qui permet une répartition plus souple des ressources.

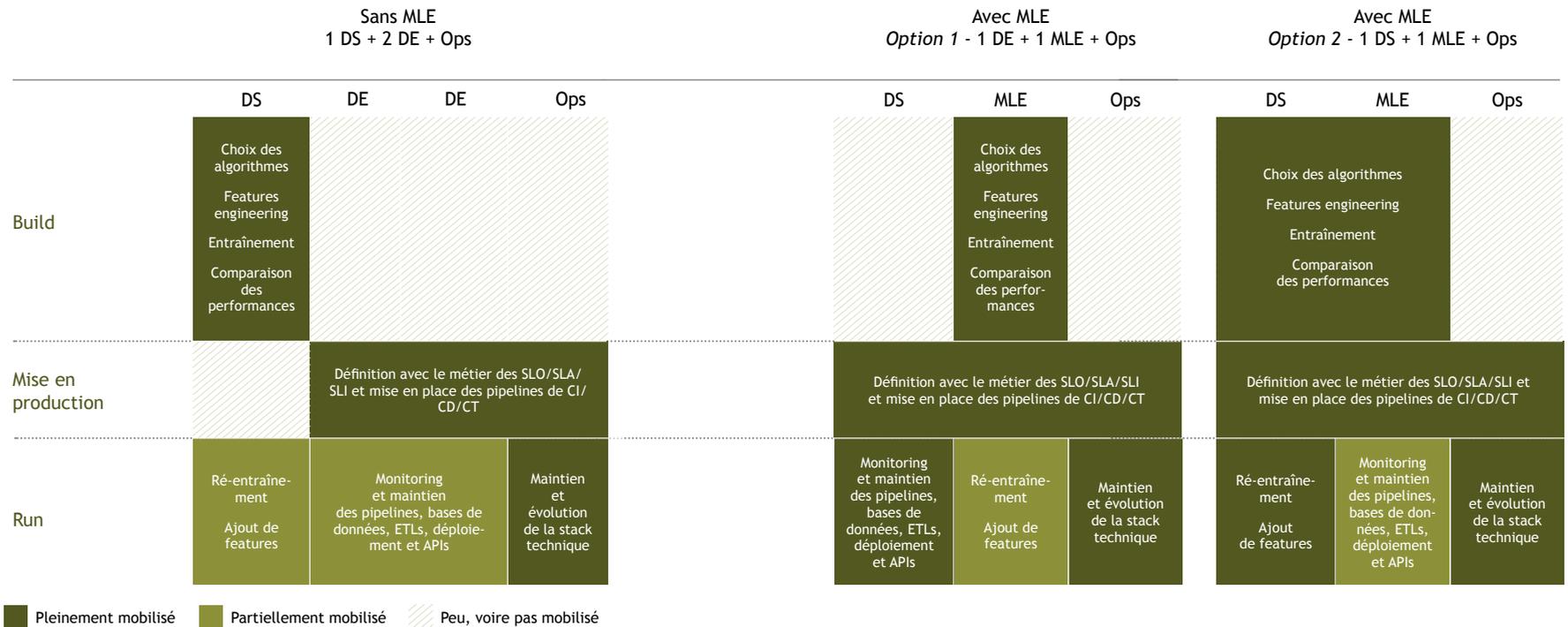


MLE - de quoi parle-t-on ? (suite)

Schéma simplifié de répartition des activités selon la configuration d'équipe

Le schéma ci-dessous permet de visualiser l'intérêt d'un profil polyvalent comme le MLE. En binôme avec un DS, il peut accélérer la phase de build sur des modèles complexes et prendre en charge le rôle d'un DE sur la mise en production et le run, si l'infrastructure n'est pas trop complexe.

En binôme avec un DE, il assure à lui seul le rôle de DS pour des modèles simples. Dans un cas plus complexe, il s'appuie sur un DE pour la mise en production et le run si l'infrastructure nécessite des compétences plus avancées.



Défi #3

Comblant le fossé culturel entre data et métiers

La plupart des frictions entre métiers et data provient de causes connues.

Frictions entre métiers et data

Symptôme

Faible utilisation des cas d'usage développés par l'équipe data

Shadow IT Data

Implication insuffisante des métiers dans le développement

Attentes démesurées des métiers

Cause

Mauvaise sélection des cas d'usage

Méconnaissance ou manque de pertinence de l'offre de services data

Manque de sponsorship

Méconnaissance du ROI*

On souffre encore du manque d'acculturation de certains métiers. Ils mésestiment souvent le temps et l'investissement nécessaires pour aller au bout de l'industrialisation. 🗨️

Karim LOUEDEC,
Group Chief Data Officer, Servier

Le manque d'acculturation des métiers aux enjeux data, de la phase d'idéation et de cadrage jusqu'à la phase d'adoption, est régulièrement un frein pour implémenter les projets data. 🗨️

Anne Laure CEBILE,
Data Science Project Manager, Danone

La poursuite de l'accompagnement des métiers en termes de sensibilisation et de formation est nécessaire pour qu'ils puissent devenir pleinement propriétaires des modèles. 🗨️

Maxime HAVEZ,
Chief Data Officer, Arkea

Les métiers et le top management détiennent une partie de la solution

Côté métier - 3 solutions pour combler le fossé

#1 -

Inscrire la data dans les grands programmes stratégiques de l'entreprise. Les cas d'usage data sont encore souvent menés par opportunisme, sans ancrage dans des projets de transformation d'envergure.

#2 -

Prendre la responsabilité du ROI* des projets data de leur côté. De la sélection jusqu'au décommissionnement d'un projet data, le métier doit être responsable du ROI. Même s'il n'a pas d'équipe data propre, il doit quand même justifier de la bonne utilisation des ressources data mises à disposition.

#3 -

Rendre le ROI* objectif (dans la mesure du possible). Le ROI des projets est rarement estimé et plus rarement encore sur les projets en run. Cela provoque des incompréhensions côté métier sur l'investissement à fournir (effort, dépenses...) autant que sur les performances à attendre d'un projet data.

Pour aller au bout du modèle d'organisation hybride, il faut que les entités métiers soient leader sur l'analyse de la création de valeur. 🗨️

Maxime HAVEZ,
Chief Data Officer, Arkea

On ne devrait pas parler de projets de data science mais de projets de transformation et de réalignement de processus métier. Cela doit devenir des projets co-pilotés avec des métiers qui s'impliquent. 🗨️

Anne Laure CEBILE,
Data Science Project Manager, Danone

Les métiers intègrent la création de valeur et le retour sur investissement (ROI) dans la présentation de leurs projets. Ces items font l'objet d'un suivi dans le cadre de notre gouvernance. 🗨️

Stanislas DRILHON,
Responsable Acculturation Data BCEF, BNP PARIBAS

*Dans ce rapport, le terme ROI s'emploie au sens large. Il inclut des indicateurs tels que le bénéfice, la marge ou le rendement, en complément du retour sur investissement.

Défi #3

Comblent le fossé culturel entre data et métiers

Les directions data détiennent l'autre partie de la solution

Côté data - Un principe de pilotage par la valeur à instaurer

Sans définition commune de la valeur, data et métiers opèrent sur des plans séparés. Pourtant :

- Seules 29% des directions data ont aligné leurs objectifs avec des métriques métiers
- En général, les directions data pilotent leur activité sur la base d'objectifs opérationnels qui leur sont propres
- Ces objectifs rendent compte uniquement des activités réalisées et des moyens engagés sans perspective nette sur l'impact métier

Exemples d'objectifs d'activité :

- Formaliser la politique de data gouvernance
- Lancer un programme d'acculturation
- Accroître le nombre de collaborateurs formés
- Augmenter la satisfaction des utilisateurs
- ...

Exemples d'objectifs de moyens :

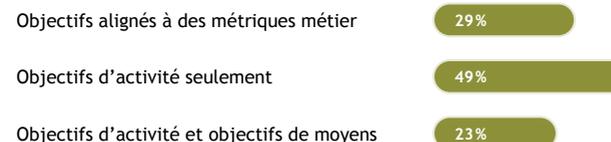
- Renforcer les effectifs / recrutements
- Installer une nouvelle capacité technologique
- Développer le réseau de contributeurs data
- Disposer de plus de données sur la plateforme
- ...

C'est le niveau de budget important qui a provoqué la mise en place d'objectifs ROI et le besoin de mesure en run. ”

Karim LOUEDEC,
Group Chief Data Officer, Servier

Recours à différents types d'objectifs pour le pilotage de l'activité data

% de répondants



Un pilotage par la valeur nécessaire pour le prochain cap de maturité

- Dépasser la phase de désillusion en sélectionnant mieux les projets IA porteurs
- Débloquer les investissements et les budgets nécessaires à l'hybridation des architectures (ex : Data Mesh)
- Convaincre les métiers de contribuer davantage au data management et à la data gouvernance
- Distinguer objectivement les bons et les mauvais arbitrages en matière de technologie, d'organisation etc.

Il faut toutefois résister à la tentation d'un pilotage excessivement bureaucratique. Le ROI doit rester un critère parmi d'autres pour ne pas paralyser l'organisation, ni générer une fausse impression de certitude.

Le ROI est fondamental mais il ne faut pas être obsédé par cette mesure. Ce que l'on cherche à faire, c'est transformer des processus existants. C'est difficile et franchement futile d'isoler la part exacte jouée par l'IA dans une telle transformation. ”

Chief Data Officer

Défi #3

Comblent le fossé culturel entre data et métiers

Un pilotage par la valeur doit s'effectuer tout le long du cycle de vie des projets

L'effort de valorisation doit s'ajuster aux questions adressées à chaque étape du cycle de vie d'un projet data et au besoin de précision. Il n'est pas nécessaire ni recommandé d'exiger systématiquement une estimation fine du ROI.

Etape du cycle de vie

Idéation et cadrage du cas d'usage	
Enjeux	<ul style="list-style-type: none"> • Filtrer et prioriser une multitude d'opportunités • Aligner le portefeuille aux enjeux stratégiques
	<ul style="list-style-type: none"> • Justifier des investissements capacitaires/coûts indirects
Méthode	<ul style="list-style-type: none"> • Critères qualitatifs d'alignement stratégique
	<ul style="list-style-type: none"> • Sizing simplifié. Ex : projets classés en tailles de T-shirt S, M, L, XL
	<ul style="list-style-type: none"> • Abaques de coûts

Niveau de Précision : Rudimentaire Moyen Fin

Build	Run
<ul style="list-style-type: none"> • Dimensionner les équipes • Valider les attentes métiers • Sécuriser le sponsorship & la contribution métier • Justifier des investissements capacitaires/coûts indirects 	<ul style="list-style-type: none"> • Dimensionner les équipes • Détecter des dérives de performance • Scale-up/down ou décommissionnement • Justifier des investissements capacitaires/coûts indirects
<ul style="list-style-type: none"> • Benchmark de projets antérieurs 	<ul style="list-style-type: none"> • Suivi du niveau de service attendu
<ul style="list-style-type: none"> • Plans de charge & plannings standards 	<ul style="list-style-type: none"> • Indicateurs d'usage et d'adoption du cas d'usage
<ul style="list-style-type: none"> • Profit&Loss/modélisation économique 	<ul style="list-style-type: none"> • Indicateurs de résultat et de performance du cas d'usage

Il est crucial pour le métier de bien clarifier aux équipes data l'impact économique attendu et les métriques associées

Impact économique possible des cas d'usage data - non exhaustif

	Inducteur	Levier
Vente/ marketing (top-line)	Base clients	Acquérir de nouveaux clients
		Fidéliser les clients existants
	Prix	Augmenter le prix de vente moyen
	Volume	Augmenter le volume de vente total
	Mixte	Augmenter la Valeur de Vie Client
Efficience opérationnelle (bottom-line)	Consomma- tion intermé- diaire	Réduire le prix d'achat unitaire d'un entrant
		Réduire les disqualités d'approvisionnement
		Réduire la consommation d'entrants
	Productivité du travail	Réduire le temps non productif
		Réduire la durée d'une tâche
		Réduire le coût horaire moyen d'un agent*
	Productivité du capital Machine & immobilier	Réduire le temps d'arrêt
		Réduire les stocks
		Réduire les défauts de production

* En transférant une tâche à un agent moins qualifié

Cas d'usage type	Calcul du gain - simplifié
Publicité ciblée	$\text{Nouveaux clients (\#)} \times \text{Valeur Vie Client (\€)}$
Prédiction du taux d'attrition	$\text{Clients retenus (\#)} \times \text{Valeur Vie Client (\€)}$
Pricing dynamique	$\text{Ventes (\#)} \times \text{hausse moyenne prix (\€)} \times \text{période (an)}$
Devis automatisé	$\text{Hausse ventes (\#)} \times \text{prix moyen (\€)} \times \text{période (an)}$
Cross-sell/up-sell	$\text{Clients (\#)} \times \text{hausse du panier moyen (\€)} \times \text{période (an)}$
Achats prédictifs	$\text{Achats (\#)} \times \text{baisse moyenne du prix (\€)} \times \text{période (an)}$
Détection de défaut sur images	$\text{Achats (\€)} \times \text{disqualités évitées (\%)}$
Reco. de paramètres de production	$\text{Achats (\€)} \times \text{réduction de volume à production constante (\#)}$
Optimisation du pas de maintenance	$\text{Réduction d'ETP (\#)} \times \text{rémunération moyenne des ETPs (\€)}$
Calcul automatique d'indicateurs	$\text{Réduction d'ETP (\#)} \times \text{rémunération moyenne des ETPs (\€)}$
Aide au diagnostique de panne	$\text{ETP (\#)} \times \text{baisse de la rémunération moyenne des ETPs (\€)}$
Maintenance prédictive	$\text{Temps gagné (h)} \times \text{Rendement horaire (\€/h)}$
Prévisions de demande	$\text{Espace libéré (\%)} \times \text{immobilisation entrepôt (\€)}$
Détection d'anomalie	$\text{Coût unitaire de fabrication (\€)} \times \text{réduction défauts (\%)}$

Défi #3

Comblent le fossé culturel entre data et métiers

En contrepartie, les équipes data et IT doivent proposer une vision claire de l'effort et des coûts à consentir

Éléments à prendre en considération dans le coût total de possession d'un cas d'usage data - non exhaustif

Inducteur	Lever	Phase		Coûts		Calcul du coût - simplifié
		Build	Run	Direct	Indirect	
Données	Achat de données externes	●	●	●	●	Prix de licence (€) x clé de répartition (%)
	Labélisation données - interne ou prestataire	●	●	●	●	Cadence de labélisation (#/h) x Labels (#) x Coût horaire moyen (€) x clé de répartition (%)
Gestion de projet	Cadrage	●	○	●	○	ETP métier, data & IT (#) x rémunération moyenne des ETPs (€)
	Gouvernance projet	●	●	●	○	ETP métier, data & IT (#) x rémunération moyenne des ETPs (€)
	Monitoring des dérives	○	●	●	○	ETP métier, data (#) x rémunération moyenne des ETPs (€)
	Monitoring de l'impact économique	○	●	●	○	ETP métier, data (#) x rémunération moyenne des ETPs (€)
Développement	Préparation des données	●	●	●	○	ETP métier, data & IT (#) x rémunération moyenne des ETPs (€)
	(Ré) entraînement du modèle	●	●	●	○	ETP métier, data & IT (#) x rémunération moyenne des ETPs (€) x fréquence (#/an)
	Sprint de développement et tests	●	●	●	○	ETP métier, data & IT (#/sprint) x rémunération moyenne des ETPs (€) x Sprints (#)
	Mise en production et tests	●	●	●	○	ETP métier, data & IT (#) x rémunération moyenne des ETPs (€)
Socle technique	Maintenance de la solution	●	●	●	○	ETP métier, data & IT (#) x rémunération moyenne des ETPs (€) x pas de maintenance
	Conso. CPU/GPU Entraînement - Cloud	●	●	●	○	Grille tarifaire de l'opérateur (€) x consommation (puissance, volume, durée...)
	Conso. CPU/GPU Prédiction - Cloud	●	●	●	○	Grille tarifaire de l'opérateur (€) x consommation (puissance, volume, durée...)
	Conso. stockage - Cloud	●	●	●	○	Grille tarifaire de l'opérateur (€) x consommation (puissance, volume, durée...)
	Conso. CPU/GPU Entraînement - On premise	●	●	○	●	Amortissement annuel (€) x clé de répartition (%)
	Conso. CPU/GPU Prédiction - On premise	●	●	○	●	Amortissement annuel (€) x clé de répartition (%)
	Conso. stockage - On premise	●	●	○	●	Amortissement annuel (€) x clé de répartition (%)
	Software/plateforme - licence	●	●	●	●	Prix de licence (€) x clé de répartition (%)
	Software/plateforme - à la consommation	●	●	●	●	Grille tarifaire de l'opérateur (€) x consommation (puissance, volume, durée...)
	Gouvernance & maintenance du socle	●	●	○	●	ETP Data & IT (#) x rémunération moyenne des ETPs (€)

Synthèse et analyse

Pour réussir, la «transformation data» doit devenir une transformation «par la data»

Les modèles hybrides sont reconnus par l'écrasante majorité des entreprises comme les mieux adaptés pour la transformation data. Toutefois, maintenir le juste équilibre entre décentralisation et pilotage transverse reste un exercice difficile, et de nombreuses entreprises se disent aujourd'hui freinées par un modèle d'organisation qui peine à aboutir.

L'un des défis majeurs consiste à mieux rapprocher la data et les métiers, souvent séparés par un fossé culturel et/ou organisationnel. À cet enjeu humain s'ajoute aussi une composante technologique. Pour rendre leur modèle réellement hybride, de nombreuses entreprises doivent réajuster les architectures qu'elles avaient construites sur une logique de centralisation; logique qu'elles cherchent à présent à dépasser. Enfin, à mesure que la data poursuit son développement dans l'entreprise, les compétences viennent à manquer. Le marché de l'emploi demeure hyperconcurrentiel depuis plusieurs années, amenant les entreprises à redoubler d'effort pour attirer, développer et retenir les talents.

Pour réussir, les entreprises doivent opérer un changement de paradigme : inscrire la data comme enabler des grands programmes de transformation stratégique. Cela est vertueux à plusieurs niveaux :

- **Organisationnel** : l'accostage à un programme stratégique fait naître un portefeuille d'usages à fort niveau de sponsorship, permettant de financer et d'opérationnaliser l'évolution vers les nouveaux modèles d'architecture
- **Attractivité sur le marché de travail** : la possibilité de travailler sur des projets stratégiques d'envergure représente une opportunité de se challenger et de rayonner, particulièrement valorisée par les candidats en phase de recrutement
- **Proximité data et métiers** tout au long du cycle de vie des usages data : en inscrivant les projets data dans les programmes stratégiques, l'entreprise facilite la mise en place d'équipes mixtes mobilisées ensemble lors des phases de build et de run du projet.

Conclusions et perspectives

Une maturité qui se maintient sur l'organisation, l'architecture, les usages et les ressources

Ce qui reste stable cette année

L'année 2021 a été marquée par l'adoption massive des modèles hybrides. Cette année, la majorité des entreprises a préféré se reposer sur l'**organisation data** existante plutôt que d'engager de nouvelles transformations profondes.

Même constat sur les **architectures & plateformes**. Beaucoup de directions data exploitent déjà un socle technique opérant, et ont cherché cette année à capitaliser sur les investissements réalisés par le passé. Les évolutions d'architecture initiées pour s'aligner sur les organisations hybrides en sont encore à un stade préliminaire.

Côté **enjeux & cas d'usage**, les entreprises se sont recentrées sur les fondamentaux et ont pour la plupart maintenu un pilotage du portefeuille axé sur des critères opérationnels, souvent éloignés des enjeux métiers.

En matière de **ressources & compétences**, les entreprises continuent de faire face à une pénurie de talents qui s'inscrit années après années comme un problème chronique.

Bien qu'en berne cette année, ces quatre axes semblent regagner de l'importance dans la feuille de route des Directions. Ils sont directement concernés par les défis de 2023 et après.

Défi #1

Faire aboutir le modèle organisationnel

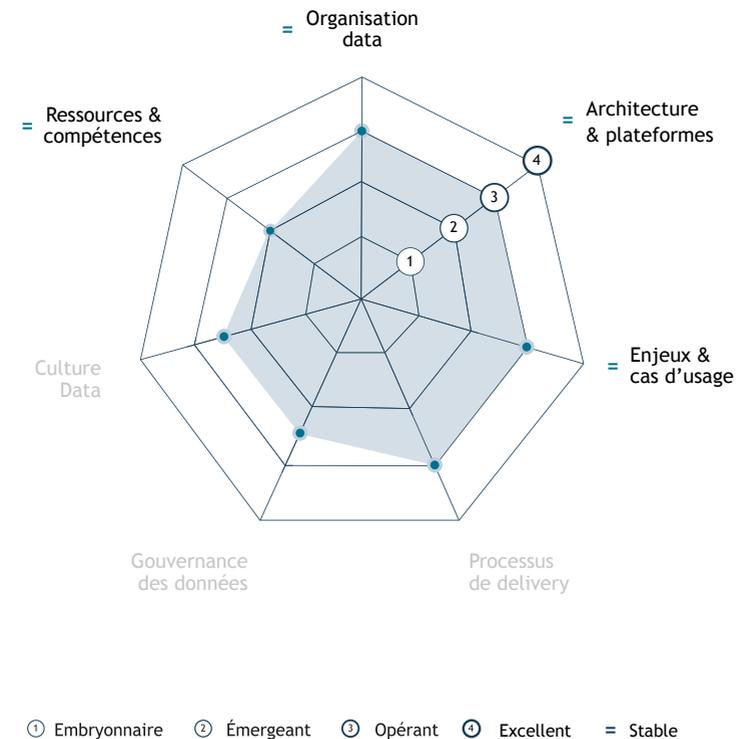
Défi #2

Sécuriser des compétences sous tension

Défi #3

Comblers le fossé culturel entre data et métiers

Evolution des scores moyens de maturité entre 2021 et 2022 sur les axes de transformation data



Une maturité qui progresse sur le delivery, la gouvernance des données et la culture data des DSI et du COMEX

Ce qui progresse cette année

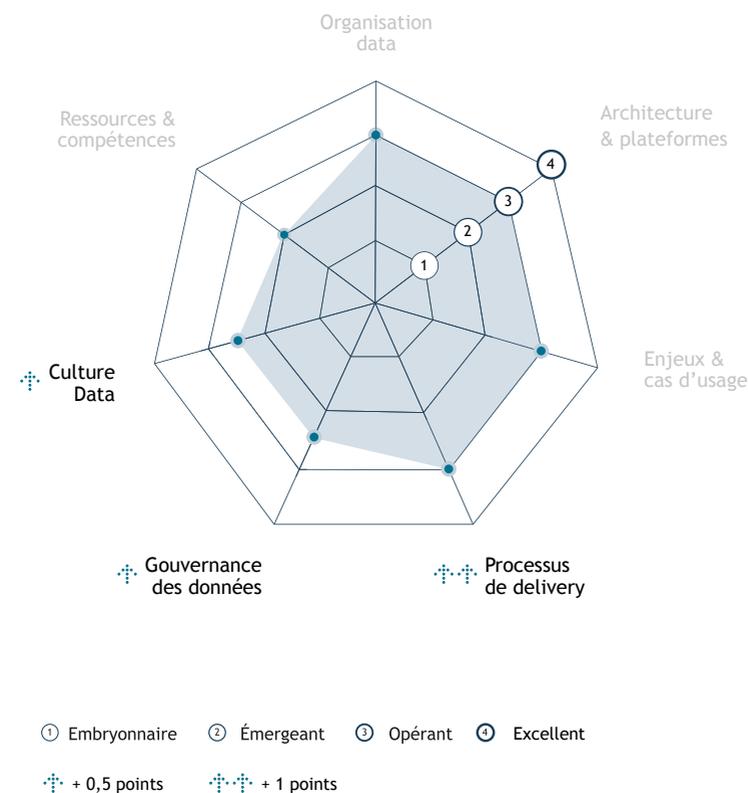
Le **processus de delivery** progresse nettement cette année alors que les entreprises fluidifient le passage en production de l'IA. Toutefois, cela ne se reflète pas clairement dans la matrice, car en franchissant un cap de maturité cette année, les entreprises ont redéfini les standards. Pour être considérées comme excellentes, elles doivent à présent tenir compte du cycle de vie complet des usages, y compris sur la phase de run, et maîtriser les enjeux organisationnels et techniques associés.

La **gouvernance de la donnée** progresse sensiblement, sous l'effet des cas d'usage de centralisation, d'accessibilité et de qualité des données, qui ont été priorités au cours de l'année. Cela dit, la contribution des métiers reste insuffisante.

La **culture data** évolue aussi, mais les progrès se cantonnent au DSI et au COMEX. Côté métier, le fossé culturel persiste.

Les axes de gouvernance et de culture seront sûrement portés en 2023 par les projets d'évolution de l'organisation hybride. Ils seront particulièrement centraux pour les entreprises adoptant les principes de Gouvernance Fédérale et de Data Product, issus du Data Mesh.

Evolution des scores moyens de maturité entre 2021 et 2022 sur les axes de transformation data



Une maturité qui progresse sur le delivery, la gouvernance des données et la culture data des DSI et du COMEX

	Enjeux & cas d'usage	Architecture & plateforme	Organisation data	Processus de delivery	Gouvernance de la donnée	Culture data	Ressources & Compétences
Excellent	La roadmap contient des cas d'usage d'envergure, souvent reliés à des programmes stratégiques menés à l'échelle de l'entreprise	Les plateformes data sont systématiquement déployées et intégrées au schéma IT global pour une mise en production à moindre coût	Le modèle opérationnel est éprouvé sur l'ensemble du périmètre	Les processus tiennent compte du cycle de vie complet des usages, y compris sur la phase de run	Les utilisateurs/ producteurs de données appliquent les standards sous la supervision des responsables Data Gouv	La data est un axe stratégique majeur. COMEX, métiers et IT soutiennent véritablement la transformation data	Les ressources sont disponibles, les RH autonomes et matures pour faire croître les effectifs
Opérant	Les cas d'usage sont cartographiés et priorisés dans une roadmap. Le ROI est un critère de sélection	Un schéma général d'architecture data est défini. Les outils technologiques sont déployés et maintenus	Les fondamentaux du modèle opérationnel sont posés (rôles, responsabilités, funding). Ils sont effectifs sur un périmètre partiel	L'industrialisation est la norme et se passe généralement de façon fluide	Les standards et politiques couvrent une partie des besoins et sont généralement peu ou mal appliqués	La data est considérée comme un patrimoine, et les directions métiers, data et DSI travaillent de concert pour le valoriser	Les effectifs nécessaires sont en place. Des écarts de compétences restent à combler par de la formation / prestation
Emergent	Des cas d'usage sont lancés de façon éparse dans l'organisation avec peu de pré-qualification	De 1ères plateformes sont déployées. Elles disposent d'environnements de développement et de production	Une ou des entités de type Datalab, Datafab, Data Office existent. Le modèle d'organisation est encore expérimental	Seuls certains usages passent en production. Ce passage reste toutefois encore exceptionnel et difficile	Les questions d'accessibilité, de sécurité/ compliance et de qualité sont gérés de façon ad hoc selon les usages	L'organisation reconnaît l'importance de la data mais les priorités et pratiques métiers, IT et data divergent	Les ressources data permettent de couvrir partiellement les besoins
Embryonnaire	Peu ou pas de besoins/ opportunités identifiés sur la data	La donnée est silotée dans les systèmes applicatifs, ou dispersée dans des documents	Les sujets data sont portés de manière désordonnée	Les cas d'usage data restent au stade d'idées ou d'expérimentation sans processus d'industrialisation	La donnée est silotée par système. L'accessibilité, la sécurité/compliance et la qualité ne sont pas pilotées	Les opportunités, enjeux et besoins propres à la data sont globalement méconnus	Pas ou peu de ressources présentes

Cases colorées : niveau moyen du marché français en 2022.

Qui sommes-nous ?



Quantmetry, The State of the Art AI company



150

Collaborateurs
et consultants-chercheurs

+ de 500

Missions IA et data

15

Prix innovation
et recherche

Un pôle éducation & upskilling, avec :

l'institut
Quantmetry



Notre promesse

Développer une IA toujours plus décisive en termes de ROI et d'impact positif global sur le business et la société, en prenant appui sur une façon de voir, de faire et de vivre le conseil en IA résolument unique sur le marché.

Quantmetry
accompagne ses clients
du cadrage stratégique
jusqu'à
l'industrialisation

Identification
des gains et définition
d'une roadmap data



Création de valeur
grâce à des solutions
IA sur-mesure



Déploiement
et généralisation des
gains à l'échelle



Les Quanters à l'initiative du baromètre



Jonathan CASSAIGNE
ASSOCIE, DIRECTEUR
DES EXPERTISES
jcassaigne@quantmetry.com

Jonathan est Associé et Directeur des expertises de Quantmetry. Il anime notre pôle Data Strategy et accompagne nos clients dans la définition de leur ambition de transformation par la data et la mise en place de leur modèle opérationnel. Il dispose de 15 ans d'expérience dans le conseil en stratégie digitale et data.



Vlad FLAMIND
MANAGER
vflamind@quantmetry.com

Fort de 10 ans d'expérience dans le conseil en stratégie et organisation, Vlad accompagne nos clients dans la construction de leur modèle opérationnel et dans la mise en œuvre concrète de programmes de Data Management et de Data Gouvernance. Il intervient aussi sur le cadrage et le pilotage d'usages data et IA au service des métiers.



Gill MORISSE
SENIOR MANAGER
gmorisse@quantmetry.com

Gill est Senior Manager au sein de la practice Data Strategy. Depuis 15 ans, il accompagne ses clients sur de grands projets de transformation par la data et sur la mise en œuvre d'intelligences artificielles de pointe. Lauréat de prix d'innovation, il intervient notamment dans le secteur de la santé, de la banque et des institutions publiques.



Carsten BONOMO
CONSULTANT
cbonomo@quantmetry.com

Carsten est Consultant en stratégie digitale dans la practice Data Strategy. Il accompagne nos clients dans la construction de leur feuille de route data. En interface avec le data office, les métiers et la DSI, il assure la cohérence des initiatives et politiques, en matière de technologie, d'organisation et de compétences.

Avec la collaboration de l'équipe commerciale Quantmetry dans la préparation et la conduite des entretiens, ainsi que de Claire Monjoin pour l'édition du présent rapport.

Merci aux entreprises et à leur dirigeant data pour la confiance qu'ils nous ont accordée, ainsi que pour la qualité de nos échanges

AG2R

APRIL

Arkea

Axa

Banque Palatine

BNP

BPCE

CCR

Clarins

CNP Assurances

Covea

Crédit Agricole

Dalkia

Danone

Decathlon

Enedis

Engie

Eureden

Europ Assistance

GAN

Groupama Gan Vie

Intermarché

Ipsen

L'Oréal

La Banque Postale

Leocare

Lolivier Assurance

LVMH

Matmut

MBDA

MGEN

Michelin

Ministère de l'Éducation

Ministère de l'Intérieur

Ministère des Affaires Étrangères

Orano Group

RCI Banque / Mobilize

Rothschild & Co

Safran

Saint Gobain

Saverglass

Servier

SMCP

Sodexo

ST Microelectronics

Stellantis

Suez

Technip

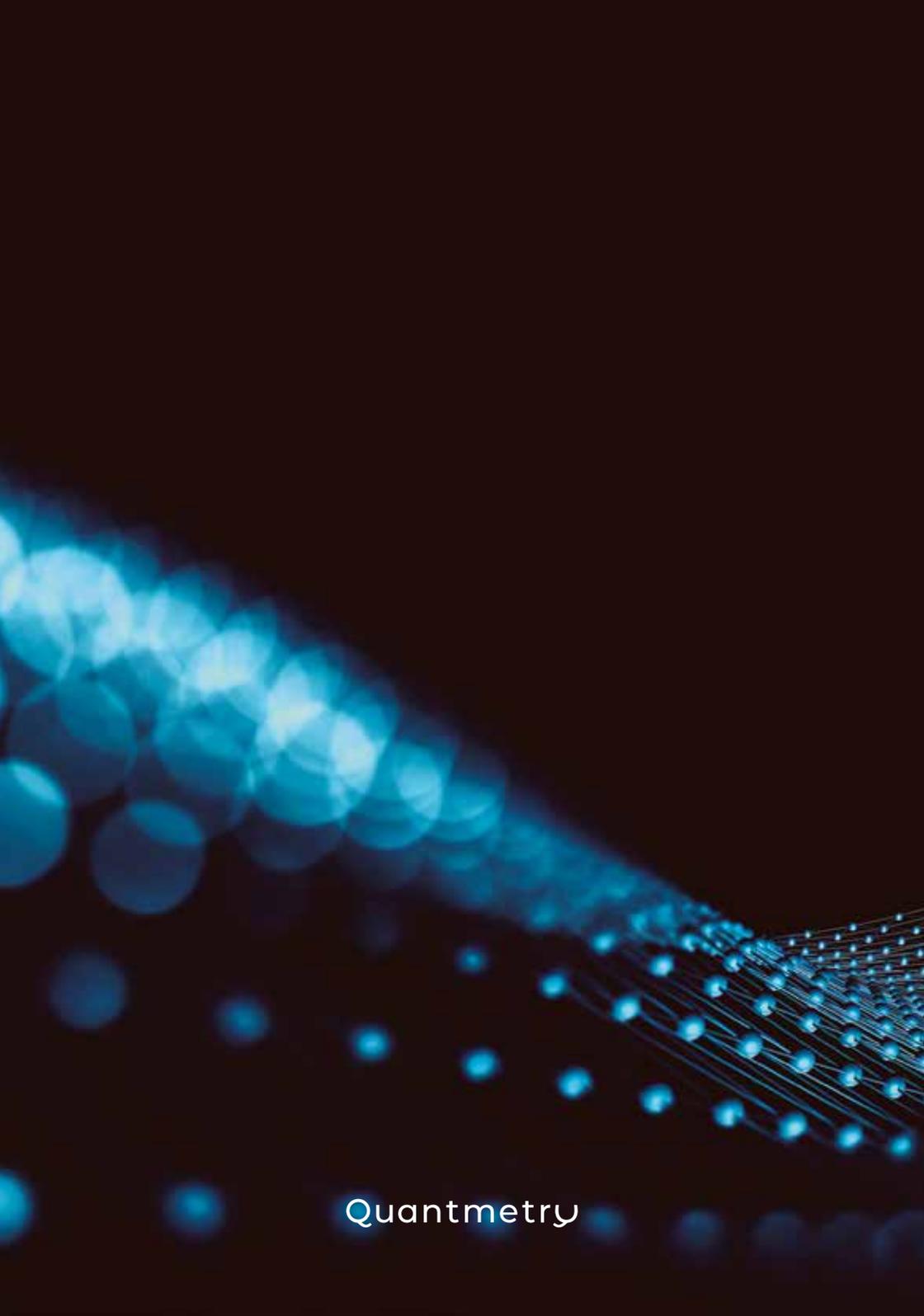
Veolia

Verlingue

Wakam

Ynsect

Yves Rocher



Quantmetry