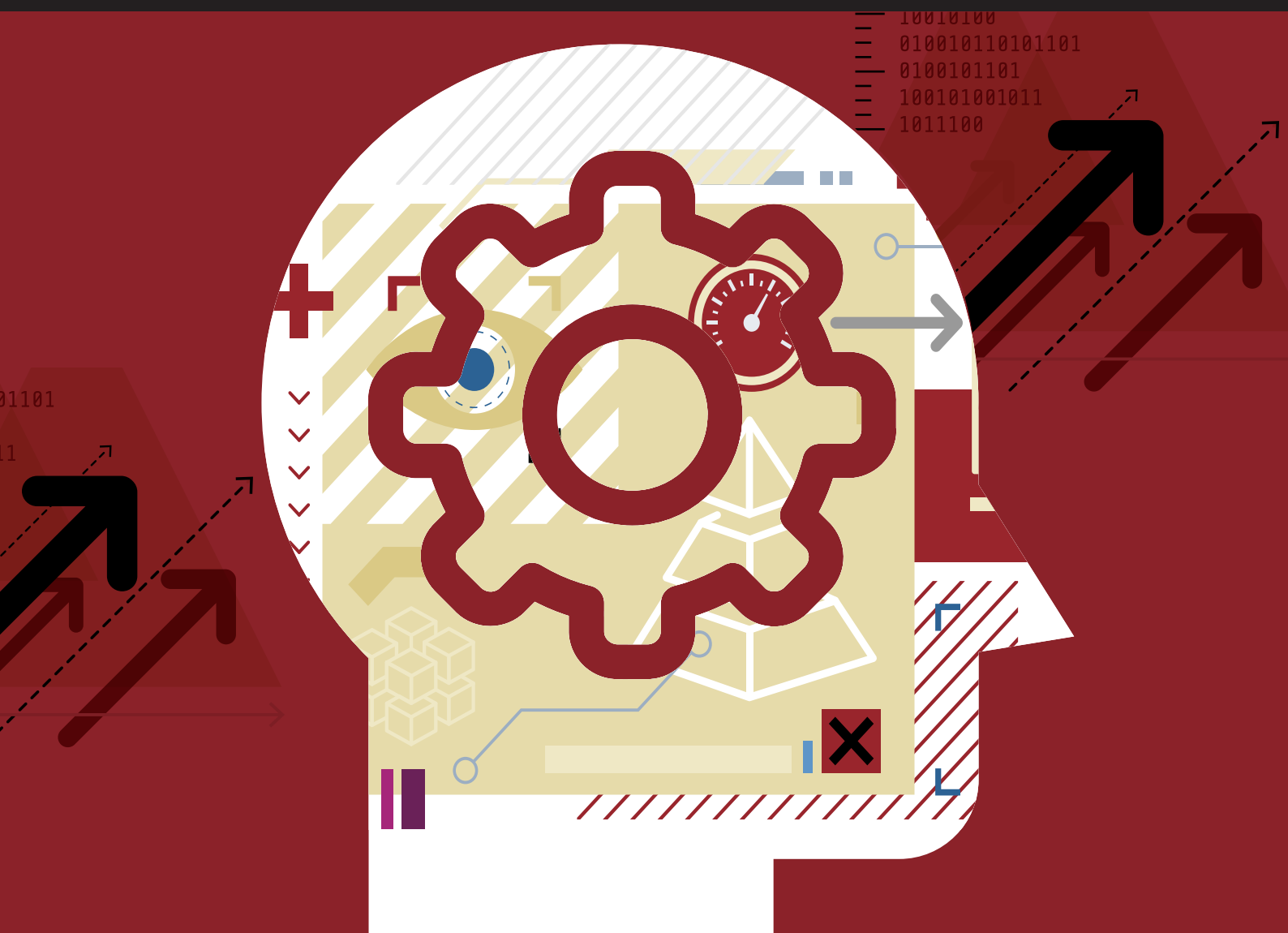


La grande accélération : perspectives des DSI sur l'IA générative



Avant-propos

« La grande accélération : perspectives des DSI sur l'IA générative » est un rapport MIT Technology Review Insights, rédigé en partenariat avec Databricks. Ce rapport s'appuie sur sept entretiens approfondis menés avec des cadres dirigeants et des experts de haut niveau en avril et mai 2023. Il s'attache à comprendre comment les leaders des technologies adoptent les outils émergents d'IA générative dans le cadre d'une stratégie d'IA à l'échelle de l'entreprise. Ce rapport tire également ses sources d'une enquête internationale menée par MIT Technology Review Insights auprès de 600 directeurs de la data et autres décideurs technologiques en mai et juin 2022. Ce rapport a été rédigé par Adam Green, édité par Teresa Elsey et produit par Nicola Crepaldi. L'étude a été menée en toute indépendance sur le plan éditorial et les avis exprimés sont ceux de MIT Technology Review Insights.

Nous souhaitons remercier les personnes suivantes de nous avoir consacré du temps et apporté leur éclairage :

Andrew Blyton, Vice-président et DSI, DuPont Water & Protection

Michael Carbin, Professeur associé, MIT, et Conseiller fondateur, MosaicML

Owen O'Connell, Vice-président senior et DSI (Services et opérations numériques de l'information), Shell

Noriko Rzonca, Chief Digital Officer, Cosmo Energy Holdings

Richard Spencer Schaefer, Directeur de l'informatique de santé, KCVA, et AI Solutions Architect, National Artificial Intelligence Institute, Département américain des Anciens combattants

Cynthia Stoddard, Vice-présidente senior et DSI, Adobe

Matei Zaharia, Cofondateur et CTO de Databricks, également Professeur associé d'informatique à l'Université de Californie à Berkeley



TABLE DES MATIÈRES

01 Résumé.....4

02 L'IA est partout en même temps6

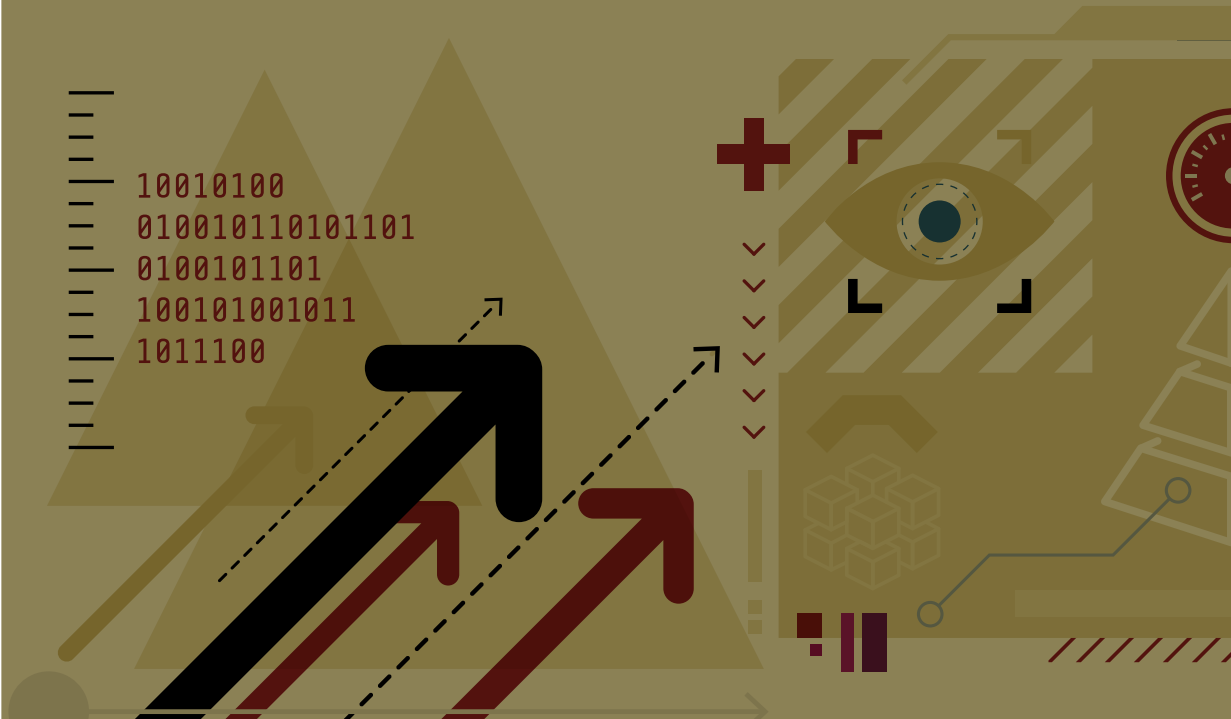
03 Bâtir pour l'IA.....10

04 Acheter ou développer ? Ouvert ou propriétaire ?.....13

05 Les inquiétudes des collaborateurs16

06 Risques et responsabilités.....19

07 Conclusion 22



01

Résumé

Avec l'émergence des outils d'IA générative accessibles au grand public dès la fin de l'année 2022 et au début de l'année 2023, le débat public autour de la puissance et du potentiel de l'IA a radicalement changé. Si l'IA générative suscite des échanges animés chez les experts depuis l'introduction de GPT 2 en 2019, les opportunités qu'elle offre aux entreprises ne sont devenues évidentes que depuis peu. Nous vivons actuellement un moment charnière dont les effets se feront sentir pendant des décennies.

L'IA générative va avoir des conséquences sur les économies et les entreprises. Selon McKinsey Global Institute, l'IA générative va injecter chaque année entre 2 600 et 4 400 milliards de dollars dans l'économie globale, accroissant de 15 à 40 % l'impact économique global de l'IA.¹ Le cabinet estime que l'IA aura automatisé la moitié du travail entre 2040 et 2060, soit dix ans plus tôt que prévu, du fait de

l'IA générative. D'après Goldman Sachs, l'IA générative sera responsable d'une augmentation de 7 % du PIB mondial. Deux tiers des emplois américains seront touchés par l'automatisation permise par l'IA.²

Les systèmes d'IA générant des textes reposent sur de grands modèles de langage (LLM), entraînés sur de vastes corpus de données. Leurs réponses ne sont pas la synthèse de leur recherche, mais le résultat de modèles mathématiques prédisant le terme suivant le plus probable.³ « La première fois que j'ai interagi avec ChatGPT, c'est son niveau de conversation qui m'a le plus frappé, » raconte Michael Carbin, Professeur associé au MIT et Conseiller fondateur de MosaicML. « Pour la toute première fois, j'ai eu le sentiment que je pouvais communiquer avec un ordinateur et qu'il interprétait ce que je disais. Selon moi, c'est la plus grande révolution qui soit depuis l'invention de l'ordinateur personnel. »

« Selon moi, c'est la plus grande révolution depuis l'invention de l'ordinateur personnel. »

Michael Carbin, Professeur associé, MIT, et Conseiller fondateur, MosaicML



Certes, l'IA était déjà considérée comme un atout stratégique bien avant l'avènement de l'IA générative. Pourtant, notre étude de 2022 avait révélé que les ambitions des DSI étaient encore modestes dans le domaine. 94 % des organisations utilisaient l'IA dans une certaine mesure, mais elles n'étaient que 14 % à envisager une généralisation de son usage « à l'échelle de l'entreprise » d'ici 2025. À l'opposé, la capacité des outils d'IA générative à démocratiser l'intelligence artificielle dans toute l'entreprise marque un tournant majeur.

Dans ce contexte, les DSI et les leaders techniques vont devoir prendre des mesures décisives : adopter pleinement l'IA générative et saisir ses opportunités pour ne pas céder de terrain à la concurrence. Ils devront aussi se positionner d'un point de vue stratégique quant à l'infrastructure des données, la propriété des modèles, la structure des équipes et la gouvernance de l'IA. Autant de choix qui auront des répercussions à long terme sur le succès de leur organisation.

Ce rapport s'intéresse aux réflexions les plus récentes des DSI d'entreprises mondialement connues, ainsi qu'à celles d'experts des secteurs public, privé et universitaire. Leurs éclairages sont mis en perspective grâce aux points de vue exprimés par 600 directeurs des données et décideurs technologiques interrogés au cours de notre étude internationale.⁴

En voici les principales conclusions :

- **L'IA générative et les LLM démocratisent l'accès à l'intelligence artificielle et ouvrent enfin la voie à une véritable généralisation à l'échelle de l'entreprise.** Stimulée par le potentiel des cas d'usage émergents, l'IA dépasse enfin le stade des projets pilote et des « îlots d'excellence » pour se généraliser en s'intégrant aux workflows organisationnels. Les équipes technologiques n'ont plus besoin de « vendre » l'IA aux fonctions métier : la demande émanant de l'entreprise exerce une traction suffisante.
- **Une véritable mine de données non structurées et dissimulées, désormais accessibles et lisibles, va devenir une nouvelle source de valeur commerciale.** Les précédentes initiatives d'IA concernaient essentiellement des cas d'usage où des données structurées étaient disponibles en abondance. En effet, collecter, annoter et synthétiser des ensembles de données hétérogènes était d'une telle complexité qu'elle rendait les initiatives d'IA à plus grande échelle inenvisageables. Aujourd'hui, la situation est différente : parce qu'elle sait mettre au jour et exploiter des données jusque-là hors d'accès, l'IA générative va être le moteur d'avancées spectaculaires dans toute l'organisation.

- **L'ère de l'IA générative doit pouvoir compter sur une infrastructure de données à la fois flexible, évolutive et efficace.** Pour soutenir ces nouvelles initiatives, les DSI et les décideurs techniques optent pour des infrastructures de données de nouvelle génération. Des approches plus sophistiquées, comme les lakehouses de données, ont le pouvoir de démocratiser l'accès aux données et à l'analytique, de renforcer la sécurité et de concilier faible coût de stockage et performance des requêtes.

- **Certaines organisations se tournent vers les technologies open source pour développer leurs propres LLM afin de protéger et rentabiliser leurs données et leur propriété intellectuelle.** Les DSI sont déjà parfaitement conscients des limites et des risques inhérents aux services tiers. Ils s'inquiètent des possibles fuites d'informations sensibles et ne veulent pas dépendre de plateformes sur lesquelles ils n'ont que peu de contrôle et de visibilité. Ils voient également des opportunités dans le développement de LLM personnalisés et comprennent la valeur des modèles de moindre envergure. Les organisations qui réussiront seront celles qui trouveront le juste équilibre stratégique entre des risques méticuleusement calculés, l'avantage compétitif et la gouvernance.

- **L'anxiété suscitée par l'automatisation ne doit pas être ignorée, mais les prédictions dystopiques sont largement exagérées.** Les outils d'IA générative peuvent déjà réaliser des tâches complexes et variées, mais les DSI et les universitaires que nous avons rencontrés n'y voient aucune menace à grande échelle. Ils entrevoient plutôt que la main d'œuvre sera globalement libérée des tâches les plus chronophages et pourra se consacrer à des domaines plus rentables en termes d'information, de stratégie et de valeur commerciale.

- **L'accélération de l'IA sera catalysée par une gouvernance unifiée et cohérente.** L'IA générative comporte des risques commerciaux et sociétaux, notamment pour la protection des secrets industriels sensibles, les infractions au droit d'auteur, le manque de fiabilité ou de transparence des résultats, et les contenus toxiques. Pour innover rapidement sans causer de dommages, ni devancer les évolutions des réglementations, les DSI doivent relever les défis de gouvernance spécifiques que pose l'IA générative, en investissant dans les processus, les technologies et les structures institutionnelles.

02

L'IA est partout en même temps

Grâce à sa flexibilité, son champ d'action et son interface conviviale basée sur le langage naturel, l'IA générative fait la preuve de ses talents dans tous les domaines, de la rédaction de textes à la production de code. Elle a le pouvoir de révolutionner les méthodes de travail dans l'ensemble des secteurs industriels et des fonctions métier. Cela laisse penser que son impact aura autant d'ampleur que celui de l'ordinateur personnel, d'Internet ou du smartphone. Elle devrait impulser la création de modèles commerciaux inédits et faire émerger de nouveaux leaders de l'industrie, au point que nous aurons bientôt oublié comment nous faisons pour travailler avant son avènement (figures 1 et 2).

L'intérêt que suscitent depuis peu la puissance et le potentiel de l'IA générative représente un tournant dans la vision que les entreprises ont de l'IA. Jusqu'à récemment, l'adoption de

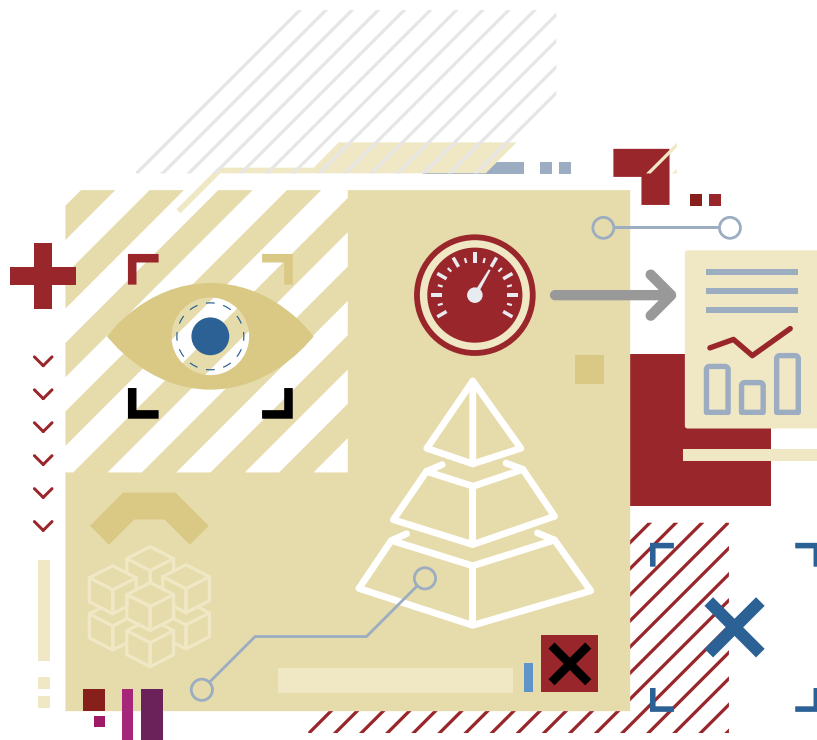
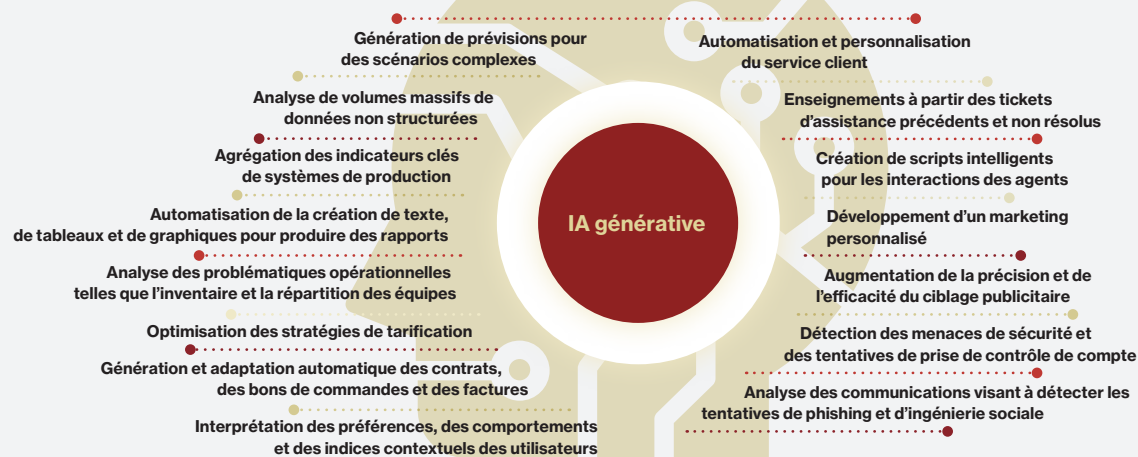


Figure 1 : Applications et cas d'usage de l'IA générative en entreprise

Ce n'est qu'une sélection des fonctions métier auxquelles l'IA générative peut être appliquée de façon convaincante.



Source : Compilé par MIT Technology Review Insights, sur la base des données issues de « Retail in the Age of Generative AI », ⁵ « The Great Unlock: Large Language Models in Manufacturing », ⁶ « Generative AI Is Everything Everywhere, All at Once », ⁷ et « Large Language Models in Media & Entertainment », ⁸ Databricks, avril-juin 2023.

l'IA était inégale d'un secteur à l'autre, et même au sein d'une entreprise. Dans notre étude de 2022, seuls 8 % des participants disaient que l'IA jouait un rôle critique dans au moins trois fonctions métier. Ce n'est que dans la finance et l'IT que plus de la moitié des participants indiquaient que l'IA constituait un aspect crucial ou significatif de la fonction.

Avant l'avènement de l'IA générative, seules quelques rares organisations avaient fait de l'IA un pilier stratégique de l'ensemble de leur activité – ou aspiraient à le faire. Certes, elles étaient déjà 94 % à utiliser l'IA d'une façon ou d'une autre, mais 14 % seulement avaient pour objectif d'exploiter « l'IA à l'échelle de l'entreprise » d'ici 2025 ; autrement dit, de rendre l'IA critique pour au moins cinq fonctions centrales (figures 3 et 4).

Mais l'IA générative a changé la nature du débat. Avec des applications dans toutes les fonctions métier, l'IA est destinée à se répandre dans toute l'entreprise. Elle s'aventure même avec confiance dans des applications créatives, un terrain jusque-là considéré comme l'apanage de l'humain. Prenons

l'exemple de Firefly d'Adobe, une famille de modèles d'IA générative créatifs qui jouent le rôle de copilote dans les workflows de création et de conception, d'après la description qu'en fait Cynthia Stoddard, Vice-présidente senior et DSI de l'entreprise. Les outils Firefly peuvent coloriser des images existantes, en créer de nouvelles, ajouter et retirer des objets d'un visuel – tout cela à partir d'instructions textuelles.

Les secteurs de l'énergie et de la chimie appliquent l'IA à des domaines auparavant inaccessibles. Le géant industriel DuPont, par exemple, avait cherché à intégrer des robots conversationnels à des interfaces s'adressant aux collaborateurs et aux consommateurs, mais leur imprécision était source de frustration. « Les LLM sont désormais en mesure d'atteindre le niveau de précision requis, bien plus rapidement, » explique Andrew Blyton, Vice-président et DSI de DuPont Water & Protection. L'entreprise utilise désormais l'IA à des fins de planification de la production, de fiabilité et de maintenance prédictive, ainsi que dans des applications d'optimisation des prix de vente.

Figure 2 : Applications et cas d'usage spécifiques

Chaque entreprise, quel que soit son secteur, trouvera ses propres applications issues de la technologie d'IA générative.



Biens de consommation et commerce de détail

- Cabines d'essayage virtuelles
- Planification de la livraison et de l'installation
- Aide à la recherche de produits en magasin
- Optimisation de la prévision de la demande et de la planification des stocks
- Génération de nouveaux concepts de produits



Production manufacturière

- Rôle de copilote expert aux côtés des techniciens
- Interactions conversationnelles avec des machines
- Services de terrain prescriptifs et proactifs
- Dépannage en langage naturel
- Évaluation du statut de garantie et de la documentation
- Compréhension des goulets d'étranglement et élaboration de stratégies de rétablissement



Médias et divertissement

- Recherche intelligente et découverte de contenu sur mesure
- Rédaction de titres et de textes engageants
- Feedback en temps réel sur la qualité des contenus
- Création de playlists personnalisées, de résumés d'actualités et de recommandations
- Création de storytelling interactif axé sur les choix des spectateurs
- Proposition d'offres et de plans d'abonnement ciblés



Services financiers

- Découverte de signaux boursiers potentiels et notification des positions vulnérables
- Accélération des décisions de souscription
- Optimisation et refonte des systèmes legacy
- Rétro-ingénierie des modèles de banque et d'assurance
- Surveillance des risques de délits financiers et de fraude
- Automatisation de la collecte à des fins de conformité
- Extraction à partir des déclarations des entreprises

Source : Compilé par MIT Technology Review Insights, sur la base des données issues de « Retail in the Age of Generative AI », ⁹ « The Great Unlock: Large Language Models in Manufacturing, » ¹⁰ « Generative AI Is Everything Everywhere, All at Once, » ¹¹ et « Large Language Models in Media & Entertainment, » ¹² Databricks, avril-juin 2023.

Figure 3 : La grande réévaluation

Avant l'IA générative, peu d'organisations avaient fait de l'IA un aspect essentiel d'une fonction métier.

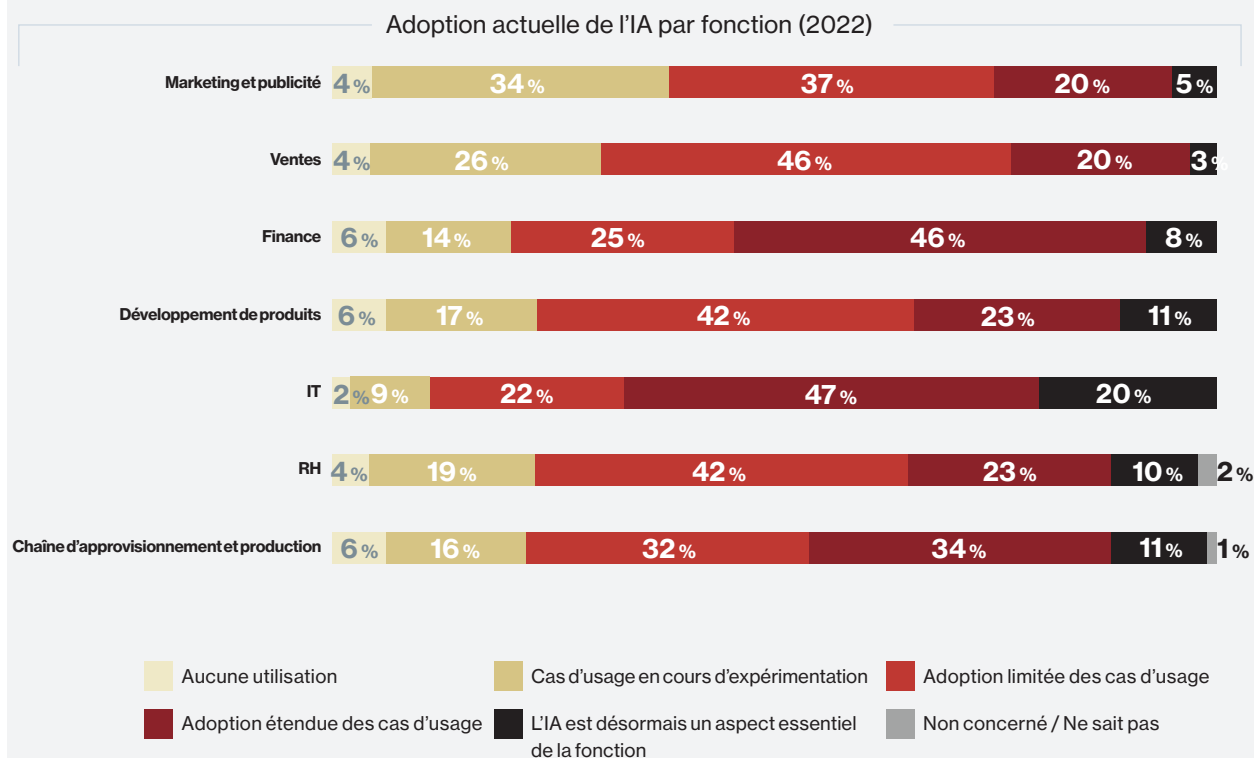
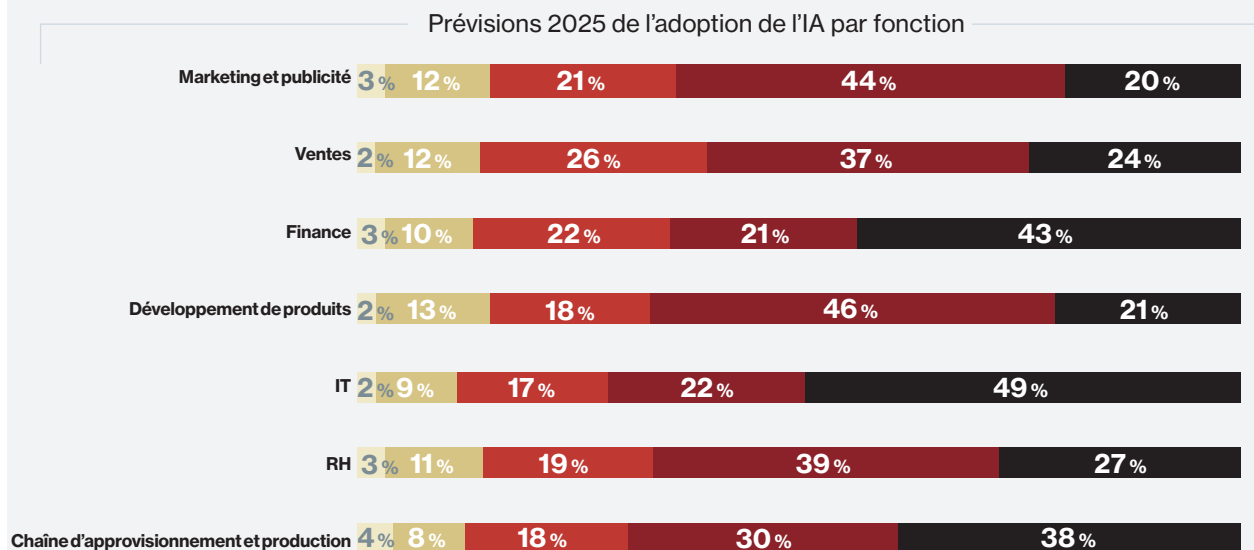


Figure 4 : Des ambitions limitées dans le domaine de l'IA

Avant l'IA générative, peu d'organisations cherchaient à faire de l'IA une capacité essentielles au sein de leurs fonctions clés à l'horizon 2025.



Les multinationales dont les actifs sont vieux de plusieurs décennies ont toujours eu les plus grandes difficultés à unifier leur infrastructure numérique. Les fusions et les acquisitions ont également fragmenté les architectures informatiques. Des documents essentiels – travaux de recherche & développement, instructions de conception d'usine, etc. – ont été perdus de vue, enfermés dans des formats de fichiers propriétaires inaccessibles. « Pourrions-nous interroger ces documents avec des LLM ? Pouvons-nous entraîner des modèles pour qu'ils nous donnent les informations qui nous échappent dans cet océan de documentation ? » s'interroge Blyton. « Pour nous, c'est un cas d'usage évident. » Les modèles de langage promettent de rendre toute leur valeur à ces données non structurées.

C'est également l'avis de Shell, le géant de l'énergie : « Nous entrevoyons déjà les bénéfices : une grande quantité de documentation stockée dans des dépôts est en train de revoir

« Les LLM ont désormais la capacité d'atteindre le degré de précision requis, bien plus rapidement. »

Andrew Blyton, Vice-président et DSI, DuPont Water & Protection

le jour, » affirme Owen O'Connell, Vice-président senior et DSI des services et opérations numériques d'information. La multinationale cherche également à harmoniser la documentation juridique, réglementaire et des ressources humaines de ses nombreuses juridictions. De nouveaux éclairages émergent même de données non structurées dans des domaines tels que le recrutement et la performance.

Grands espoirs pour la santé

La santé et la médecine, deux disciplines profondément centrées sur l'humain, sont elles aussi des terrains d'essais pour l'IA. Dans les laboratoires, les outils basés sur l'IA ont fait leur preuve en prédisant des structures de protéines, en participant à la découverte de nouveaux médicaments et en traçant la propagation des épidémies, dont la pandémie de Covid-19. L'IA générative s'annonce désormais comme un puissant assistant pour le personnel qui est en première ligne. Les outils de traitement du langage naturel savent transcrire et synthétiser des notes médicales, tandis que les robots conversationnels peuvent être entraînés à répondre aux questions médicales.

Richard Spencer Schaefer est Directeur de l'informatique de santé au sein du Centre médical des Vétérans de Kansas City, mais aussi AI Solutions Architect au sein de l'Institut national d'intelligence artificielle des Anciens combattants. Le département a récemment validé un modèle qui identifie le risque qu'un patient admis à l'hôpital soit transféré dans une unité de soins intensifs dans les 24 heures. « Nous avons constaté une amélioration considérable de la précision et de la capacité

de prédiction de ce modèle d'IA, au point qu'il pourrait entraîner une réduction significative de la mortalité, » explique-t-il.

Schaefer est également enthousiaste quant à la capacité de l'IA à soulager le personnel soignant, tout en gagnant sa confiance via des améliorations modestes mais concrètes. « Dans le domaine de la santé, poursuit-il, je pense que ce sont des améliorations et des automatisations très spécifiques de processus qui vont stimuler l'adoption de l'IA. » Le Département des Anciens combattants, par exemple, travaille actuellement sur un projet d'IA visant à réduire les alarmes inutiles dans ses centres médicaux. Des modèles de machine learning sont capables de définir des seuils d'intervention de façon dynamique et spécifique à chaque patient, bien plus utiles que les critères statiques et les traditionnels seuils des systèmes d'alarme. « Pour obtenir des résultats, conclut Schaefer, nous devons gagner la confiance des soignants et les aider à comprendre ce dont l'IA est réellement capable. »

03

Développer pour l'IA

Les applications d'IA reposent sur une solide infrastructure de données qui permet la collecte, le stockage et l'analyse de ses vastes dépôts.

Avant même que les applications commerciales de l'IA générative ne se manifestent à la fin de

2022, 70 % des participants de notre étude considéraient qu'une plateforme de données unifiée, dédiée à l'analytique et à l'IA, était essentielle (figure 5).

L'infrastructure et l'architecture des données englobent l'infrastructure logicielle et réseau, très souvent en cloud ou en cloud hybride, et du matériel, sous la forme de GPU hautement performants. Les entreprises ont besoin d'une infrastructure capable de maximiser la valeur des données sans compromettre leur sécurité, en particulier à une époque où les règles encadrant la protection des données et l'IA se multiplient. Pour véritablement démocratiser l'IA, l'infrastructure doit offrir une interface simple permettant aux utilisateurs d'interroger les données et d'exécuter des tâches complexes en langage naturel. « L'architecture évolue de manière à soutenir la démocratisation de l'analytique, » explique Schaefer.

« Nous avons agrégé des données sur de nombreuses technologies différentes au fil des années. Il nous apparaît aujourd'hui que c'est le lakehouse qui présente le meilleur rapport performance-coût dès le départ. »

Andrew Blyton, Vice-président et DSI,
DuPont Water & Protection

Les infrastructures de type lakehouse sont devenues populaires. Elles sont le fruit de l'hybridation de deux approches existantes, les data warehouses et les data lakes. Les data warehouses sont apparus dans les années 1980 pour systématiser la business intelligence et le reporting d'entreprise. Ils n'offrent toutefois pas de services en temps réel et ne permettent que du traitement par lot. Surtout, ils sont incapables de gérer les formats de données émergents et non traditionnels. Les data lakes, appréciés pour leur meilleure prise en charge des tâches d'IA et de data science, ont émergé plus récemment. Mais ils restent complexes à mettre sur pied, ils souffrent de lenteurs et les contrôles de la qualité de leurs données sont médiocres. Le lakehouse combine les avantages des deux approches. Il propose une architecture ouverte possédant à la fois la flexibilité et l'échelle des data lakes, ainsi que les possibilités de gestion et de contrôle qualité des warehouses.

Le lakehouse a notamment séduit le Département des Anciens combattants parce qu'il minimise les déplacements



Contrôler la vanne des données

L'entraînement du modèle GPT-3 d'OpenAI a coûté plus de 4 millions de dollars, et celui du modèle LLaMA de Meta plus de 2,4 millions de dollars, mobilisant 1 million d'heures de GPU.¹³ L'entraînement d'un modèle d'IA est énergivore : on parle d'une consommation d'électricité supérieure à celle de 100 foyers américains sur une année entière.¹⁴ Celui de GPT-3 a consommé 1 287 MWh d'électricité et émis plus de 550 tonnes de dioxyde de carbone.¹⁵ « Ces modèles comptent parmi les plus gros consommateurs de calcul, » souligne Owen O'Connell de Shell.

Les grands modèles coûtent cher à entraîner et à exploiter : le prix de chaque interaction est corrélé à la taille du modèle, comme l'explique Michael Carbin (MIT et MosaicML). Nous estimons qu'OpenAI dépense 40 millions de dollars chaque mois pour traiter les requêtes saisies par les utilisateurs, tandis que le robot conversationnel de Bing (Microsoft) utiliserait une infrastructure d'une valeur de 4 milliards de dollars pour servir sa base d'utilisateurs potentiels.¹⁶ Les modèles plus petits et plus efficaces (voir chapitre suivant) pourraient constituer une alternative prometteuse.

Les entreprises doivent être plus attentives à l'efficacité de leurs activités, tandis que les start-ups et les chercheurs inventent de nouvelles manières d'améliorer celle des modèles d'IA. « Face à l'augmentation des coûts, nous recrutons des ingénieurs qui savent les optimiser et les contrôler, » confie Andrew Blyton de DuPont. « Dans certains cas, nous devenons beaucoup plus efficaces sur le plan des requêtes ou du calcul parce que nous recrutons des experts chargés de l'optimisation. » DuPont suit et analyse également les coûts du cloud computing à un niveau très détaillé. « Nous avons recruté des talents et nous comptons également sur nos fournisseurs pour tout optimiser. Une mauvaise approche pourrait avoir de lourdes conséquences, à la fois en termes de temps et de performance, » explique A. Blyton.

de données, toujours source de risque de confidentialité et de gouvernance. « Nous sommes confrontés à un défi majeur concernant le déplacement des données. Nous consacrons d'importantes ressources et une grande expertise à le relever, » explique Schaefer.

DuPont a investi dans plusieurs infrastructures technologiques et le data lakehouse lui semble prometteur. « Nous avons agrégé des données sur de nombreuses technologies différentes au fil des années, déclare Blyton. Il nous apparaît aujourd'hui que c'est le lakehouse qui présente le meilleur rapport performance-coût dès le départ. Voilà pourquoi nous avons intensifié nos investissements pour l'adopter à grande échelle. » Pour DuPont, le data lakehouse offre une visibilité sur un portefeuille organisationnel complexe qui couvre des années de fusions-acquisitions et de dette technique. « C'est avant tout pour interpréter les données de nos activités que nous avons investi dans le data lakehouse, » explique Blyton. « Comment faire pour rassembler différents systèmes de planification des ressources d'entreprise (ERP), les fusionner et donner aux équipes un accès en quasi-temps réel à des informations qu'il fallait précédemment rechercher manuellement ? »

Le lakehouse masque la complexité pour permettre aux utilisateurs d'accomplir des tâches avancées, quel que soit leur niveau de compétence technique. Shell a développé une « couche entreprise » avec laquelle les utilisateurs peuvent interagir de façon dynamique. « Auparavant, il fallait accéder à des data stores, extraire les données, les nettoyer et leur appliquer plusieurs processus de transformation, » se souvient O'Connell. Avec le lakehouse, les utilisateurs exécutent des charges de travail, réalisent des analyses historiques et des recherches de tendance en toute autonomie, tandis que le cloud offre toute l'élasticité dont nous avons besoin en termes de calcul.

Le data lakehouse traite aussi bien les données structurées que non structurées, ajoute-t-il. « Les bases de données classiques sont conçues pour un seul type de données. » Le lakehouse « nous permet d'agir beaucoup plus rapidement. Nos données arrivent dans une architecture commune accessible à toutes les équipes qui peuvent en extraire de la valeur commerciale. » Ces capacités sont essentielles pour une entreprise autant orientée données que Shell. Dans le seul domaine de la maintenance prédictive, l'entreprise gère 17 000 modèles et 4 000 milliards de lignes de données. S'y ajoutent chaque jour 20 milliards de lignes supplémentaires issues de 3 millions de capteurs répartis sur ses sites dans le monde entier.

Ceux qui optent pour le data lakehouse mettent cette technologie au cœur d'une stratégie de données qui facilite l'émergence de cas d'usage concrets et impactants basés sur les données et l'IA. Comme le rappelle Schaefer, « la démocratisation des données est un processus assez coûteux, et nous manquons traditionnellement de ressources dans ce domaine. Mais nous entrons à grande vitesse dans une ère où les modèles transactionnels et les outils logiciels des lakehouses et des data lakes hybrides deviennent enfin performants ensemble. Dans le domaine de la santé en particulier, je pense que nous allons observer une adoption rapide d'outils dédiés à l'amélioration de l'efficacité

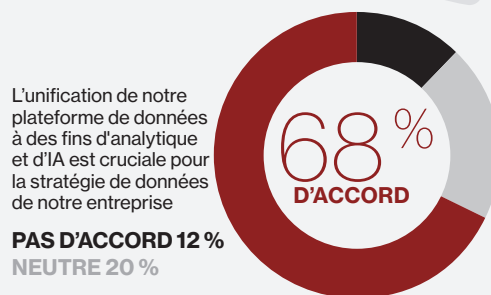
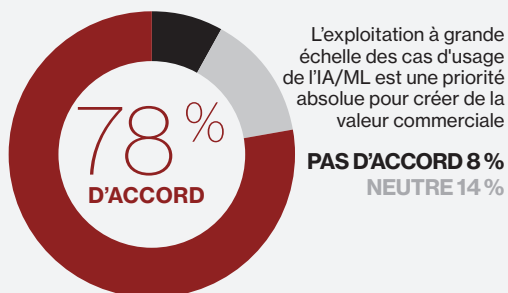
opérationnelle. Je vois ainsi d'innombrables opportunités d'améliorer nos pratiques et d'automatiser des processus. »

Quant à Blyton, il ajoute : « Chez DuPont, nous utilisons le lakehouse et des outils comme le data lake pour développer une base robuste à laquelle appliquer des algorithmes de ML. Nous élargissons la portée de la maintenance prédictive en utilisant des algorithmes avec des capteurs tiers. Cela nous permet de déterminer à quel moment une machine va tomber en panne, avant même que la défaillance ne survienne, et non après. Nous parlons de cas d'usage concrets, convaincants, chiffrés : nos investissements produisent un retour tangible et mesurable. »

« Les bases de données classiques sont conçues pour un seul type de données. Le lakehouse nous permet d'agir beaucoup plus rapidement. »

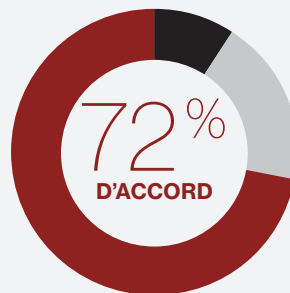
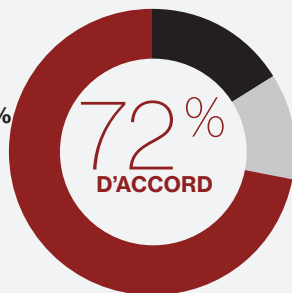
Owen O'Connell, Vice-président senior et DSI (services et opérations numériques d'information), Shell

Graphique 5 : Les priorités des dirigeants dans le domaine de l'IA



Nous privilégions une approche multicloud comme base flexible pour l'IA/ML

Opinion	Pourcentage
D'ACCORD	72 %
PAS D'ACCORD	12 %
NEUTRE	16 %



Les problèmes de données sont le principal facteur de risque pour nos objectifs d'IA/ML

Opinion	Pourcentage
D'ACCORD	72 %
PAS D'ACCORD	9 %
NEUTRE	19 %

04 Acheter ou développer ? Ouvert ou propriétaire ?

Dans le cadre de leurs réflexions sur les capacités des plateformes tierces d'IA générative, les DSI et les équipes dirigeantes réévaluent leurs idées reçues sur les questions de propriété, de partenariat et de contrôle.

Il est peu probable que la surexploitation d'une plateforme d'IA généraliste puisse conférer un quelconque avantage compétitif. Comme le dit Carbin, « Si un problème particulier vous préoccupe vraiment ou que vous avez l'intention de développer un système fondamental pour votre entreprise, il faut savoir qui possède votre propriété intellectuelle. » DuPont est « une entreprise de science et d'innovation » ajoute Blyton, « il est essentiel que nos LLM restent internes à notre organisation pour protéger et sécuriser notre propriété intellectuelle. C'est un impératif absolu. »

L'éventualité d'un risque concurrentiel inquiète. « Il peut être risqué de s'appuyer sur un modèle existant, parce que les données que vous y injectez pourraient être utilisées par une entreprise pour faire concurrence à vos produits cœur de métier, » souligne Carbin. D'autre part, les utilisateurs manquent de visibilité sur les données, les pondérations et les algorithmes qui alimentent les modèles fermés. Sans compter que le produit lui-même et ses données d'entraînement peuvent changer à tout moment. C'est un sujet particulièrement crucial dans la R&D scientifique où la reproductibilité est essentielle.¹⁷

Certains DSI prennent des mesures pour limiter l'utilisation de plateformes d'IA générative externes au sein de l'entreprise. Samsung a banni ChatGPT quand des employés s'en sont servis pour travailler sur du code sensible sur le plan commercial. Bien d'autres entreprises, dont JP Morgan Chase, Amazon et Verizon, ont également mis en place des

« Si un problème particulier vous préoccupe vraiment ou que vous avez l'intention de développer un système 'cœur de métier' pour votre entreprise, il faut savoir qui possède votre propriété intellectuelle. »

Michael Carbin, Professeur associé, MIT, et Conseiller fondateur, MosaicML

restrictions, voire des interdictions. « Nous ne pouvons pas laisser des éléments comme la stratégie d'entreprise de Shell passer par ChatGPT, » résume O'Connell. Et comme les LLM exploitent l'intégralité de l'univers connecté, les entreprises devraient, selon lui, faire preuve de davantage de prudence à l'avenir quant aux données qu'elles mettent en ligne. « Elles prennent conscience qu'un inconnu pourrait extraire une grande valeur de leurs données. »

Le manque d'exactitude et de fiabilité des résultats est une autre source d'inquiétude. Les plus grands LLM sont, du fait de leur taille, entachés de fausses informations glanées en ligne. Et cela vient étayer les arguments en faveur d'approches plus ciblées selon Matei Zaharia, cofondateur et CTO de Databricks, également Professeur associé d'informatique à l'Université de Californie à Berkeley. « Si vos projets concernent un domaine plus spécifique, explique-t-il, vous allez ainsi éviter toutes les scories et informations indésirables venues du web. »

Mais les entreprises ne peuvent pas créer leur propre version de ces modèles gigantesques. L'échelle et le coût d'un tel projet met ce type de tâche informatique hors de portée des entreprises, à l'exception des plus grandes : OpenAI dit avoir utilisé 10 000 GPU pour entraîner ChatGPT.¹⁸ À l'heure actuelle, l'élaboration de modèles de grande échelle est réservée aux entreprises technologiques les mieux armées.

Par contre, les modèles plus petits offrent une alternative viable. « Je pense que nous allons passer de 'Il me faut 500 milliards de paramètres dans un modèle' à 'Il me faut peut-être 7, 10, 30, 50 milliards de paramètres sur les données que j'ai réellement', » rappelle Carbin. « Pour réduire la complexité, il faut renoncer à l'idée d'un modèle généraliste détenteur de toute la connaissance humaine. Il est préférable d'opter pour une connaissance spécialisée de très haute qualité, parce que c'est de cela dont les collaborateurs ont réellement besoin. »

Heureusement, plus petit ne veut pas dire moins puissant. Les modèles génératifs ont été finement ajustés pour des domaines qui nécessitent moins de données, comme le montre par exemple BERT et ses variantes : BioBERT pour le contenu biomédical, Legal-BERT pour le contenu juridique et CamemBERT... pour le contenu en français (cela ne s'invente pas).¹⁹ Pour prendre en charge des cas d'usage précis, les organisations peuvent privilégier la spécificité à la connaissance généraliste. « En réalité, les gens sont en quête de modèles qui maîtrisent leur domaine, » affirme Carbin. « Une fois que vous avez pris ce virage, vous comprenez vite qu'il existe d'autres manières de fonctionner et de réussir. »

Dolly : Le LLM conversationnel à 30 \$

En mars 2023, Databricks a publié Dolly, un LLM open source visant à démocratiser la puissance de l'IA générative.²⁰

Entraîné pour moins de 30 \$, il affiche une capacité conversationnelle comparable à celle de ChatGPT. Autrement dit, il est capable de suivre les instructions de l'utilisateur. Il repose sur l'outil LLaMA de Meta et a été affiné à l'aide d'entrées de haute qualité obtenues par crowdsourcing auprès des collaborateurs de Databricks. Dolly, qui doit son nom au premier mammifère cloné au monde, ne compte que 6 milliards de paramètres, soit moins de 3,5 % des 175 milliards de paramètres utilisés dans le GPT-3.

« Les modèles très haut de gamme, qui comptent des centaines de milliards de paramètres, peuvent coûter très cher à entraîner et à exécuter, » souligne Matei Zaharia de Databricks. En raison de ces coûts, de nombreuses entreprises s'appuient sur des outils propriétaires comme ChatGPT ou dépensent des sommes et des ressources considérables pour développer le leur. Comme il le rappelle, même si des gains d'efficacité en termes de matériel et de logiciel vont réduire les coûts au fil du temps, le développement de modèles comptant des milliers de milliards de paramètres va rester hors de portée de la plupart des acteurs.

Pour fournir une alternative aux outils d'IA proposés par de très grandes entreprises, Databricks a cherché à développer un modèle conversationnel qui utiliserait du code et des données open source, mais permettrait également une utilisation commerciale. « Ce que nous voulions savoir avant tout, explique Zaharia, c'était combien coûterait de développer un modèle en s'appuyant uniquement sur des ensembles de données publics et disponibles, du code open source et des données d'entraînement elles aussi open source, afin de produire un modèle utilisable par tous. »

Databricks a ouvert le code d'entraînement, l'ensemble de données et les pondérations de Dolly. Dolly 2.0, lancé en avril 2023, est le premier LLM open source utilisable sous licence pour un usage commercial.²¹ Il permet aux entreprises de combiner leurs données avec celles de Databricks pour développer des applications personnalisées.

En donnant la possibilité de développer et personnaliser à peu de frais leurs propres outils, des plateformes comme Dolly peuvent démocratiser l'accès à l'IA générative.

« Tous les grands modèles que proposent les fournisseurs tiers sont entraînés sur les données du Web. Mais il existe dans votre entreprise d'innombrables concepts et données que ces modèles ignorent totalement. »

Matei Zaharia, CTO de Databricks, également Professeur associé d'informatique à l'Université de Californie à Berkeley

« Les gens voient de plus en plus les données comme une réserve compétitive, » développe Zaharia. « Des exemples comme BloombergGPT [un LLM spécialement conçu pour la finance] signalent clairement que les entreprises réfléchissent à l'utilisation qu'elles peuvent faire de leurs données et cherchent à commercialiser leurs propres modèles. »

« Les entreprises vont élargir et personnaliser ces modèles à partir de leurs données internes, dans le but de les intégrer à leurs propres applications, dans l'intérêt de leurs activités, » anticipe Zaharia. « Tous les grands modèles que proposent les fournisseurs tiers sont entraînés sur les données du Web. Mais il existe au sein de votre entreprise d'innombrables concepts et données que ces modèles ignorent totalement. Et le plus intéressant, c'est que le modèle n'a pas besoin d'une quantité considérable de données supplémentaires ou d'un très long entraînement pour acquérir des connaissances sur un nouveau domaine, » explique-t-il.

Les modèles open source plus modestes, comme le LLaMA de Meta, peuvent rivaliser avec les performances des grands modèles et permettre aux professionnels d'innover, d'échanger et de collaborer. Une équipe a développé un LLM en utilisant les pondérations de LLaMA pour un coût inférieur à 600 \$, alors que l'entraînement de GPT-4 a coûté 100 millions de dollars. Le modèle, baptisé Alpaca, est aussi performant que l'original sur certaines tâches.²² Et grâce à la transparence supérieure de l'open source, les chercheurs et les utilisateurs identifient plus facilement les biais et les défauts de ces modèles.

« L'essentiel de cette technologie est à la portée de nombreuses organisations, » déclare Carbin. « Pas seulement de celles du calibre d'OpenAI, Google ou Microsoft, mais aussi des entreprises de taille intermédiaire, et même des start-ups. »



05

Les inquiétudes des collaborateurs

Du téléphone à l'ordinateur, chaque bond en avant dans les technologies suscite des inquiétudes et est perçu comme une menace pour les emplois, le talent et les compétences des humains. Mais l'IA générative semble marquer un tournant plus grand encore, du fait de l'immense éventail de tâches qu'elle permet d'automatiser. Pour la première fois, des secteurs techniques de pointe et des domaines créatifs font partie des plus propices à l'automatisation.

D'après une analyse d'Accenture, 40 % des heures de travail, tous secteurs confondus, pourraient être automatisées ou enrichies par l'IA générative. Les secteurs qui affichent le plus grand potentiel sont la banque, l'assurance, les marchés de capitaux et le développement logiciel.²³ Le cabinet McKinsey prévoit que l'IA générative et les technologies connexes pourraient automatiser des activités qui représentent actuellement 60 à 70 % du temps de travail, soit 50 % de plus qu'avant l'avènement de l'IA générative.²⁴

Selon Goldman Sachs, les deux tiers des emplois américains seront concernés par l'automatisation basée sur l'IA. Mais cela ne doit pas nécessairement se traduire par des suppressions de postes massives : « La plupart des emplois ne sont que partiellement exposés. Ils sont donc davantage destinés à être complétés que remplacés par l'IA. » Le cabinet rappelle également que les pertes d'emplois liées à la technologie sont généralement compensées (et même plus) par la multiplication de postes technologiques : « Plus de 85 % de la croissance du nombre d'emplois de ces 80 dernières années s'explique par des créations de postes liées à la technologie ». ²⁵

Les DSI et les universitaires interrogés affichent leur optimisme concernant les effets de l'IA générative sur les entreprises et la société dans son ensemble. Ils pensent que l'IA peut soutenir des secteurs où la main d'œuvre est rare, comme la santé, et que l'expertise humaine va rester fondamentale. « Les gens ont

peur de perdre leur emploi à cause de l'IA, mais ce n'est pas le premier sujet d'inquiétude dans la santé, » nous dit Schaefer. « Les technologies que nous mettons en place permettent aux médecins de prendre part au développement de l'IA. Et ce niveau de validation est, selon moi, le gage d'une confiance accrue dans les modèles que nous élaborons. » Les humains restent indispensables pour annoter, sélectionner, exposer les modèles aux ensembles de données et garantir un usage responsable de l'IA. « Nous avons l'opportunité de libérer l'IA et de la mettre au service de nos recherches, parce que nous apprenons énormément les uns des autres en ce moment, » ajoute O'Connell.

L'IA générative est également un catalyseur d'autonomie. Elle démocratise en effet l'accès à des capacités techniques jusque-là réservées à une petite minorité de la main d'œuvre.

« Les technologies que nous mettons en place permettent aux médecins de prendre part au développement de l'IA. Ce niveau de validation est, selon moi, le gage d'une confiance accrue dans les modèles que nous élaborons. »

Richard Spencer Schaefer,
Directeur de l'informatique de santé,
Centre médical des Vétérans
de Kansas City

Les entreprises prennent déjà des mesures pour encourager l'engagement. « Nous avons constitué en interne des communautés de professionnels qui n'ont pas nécessairement un profil informatique typique, » explique Blyton. « Nous voulons donner à nos utilisateurs métier la possibilité d'élaborer leurs propres tableaux de bord et d'extraire eux-mêmes des insights à partir des données. » Cette démocratisation de l'accès aux données figurait d'ailleurs parmi les principales approches utilisées par les entreprises pour générer des bénéfices tangibles à partir de l'IA en 2022, selon notre enquête (figure 6).

Blyton prédit que c'est au sein des équipes que les idées d'utilisation de l'IA vont émerger, marquant le début d'une ère davantage axée sur le self service et l'autonomie entrepreneuriale au sein des organisations. Selon O'Connell, les dirigeants des unités commerciales, informés des possibilités de l'IA par l'intense couverture médiatique de ces derniers mois, réclament déjà des applications d'IA sans que les équipes technologiques n'aient à « vendre » l'IA à l'organisation.

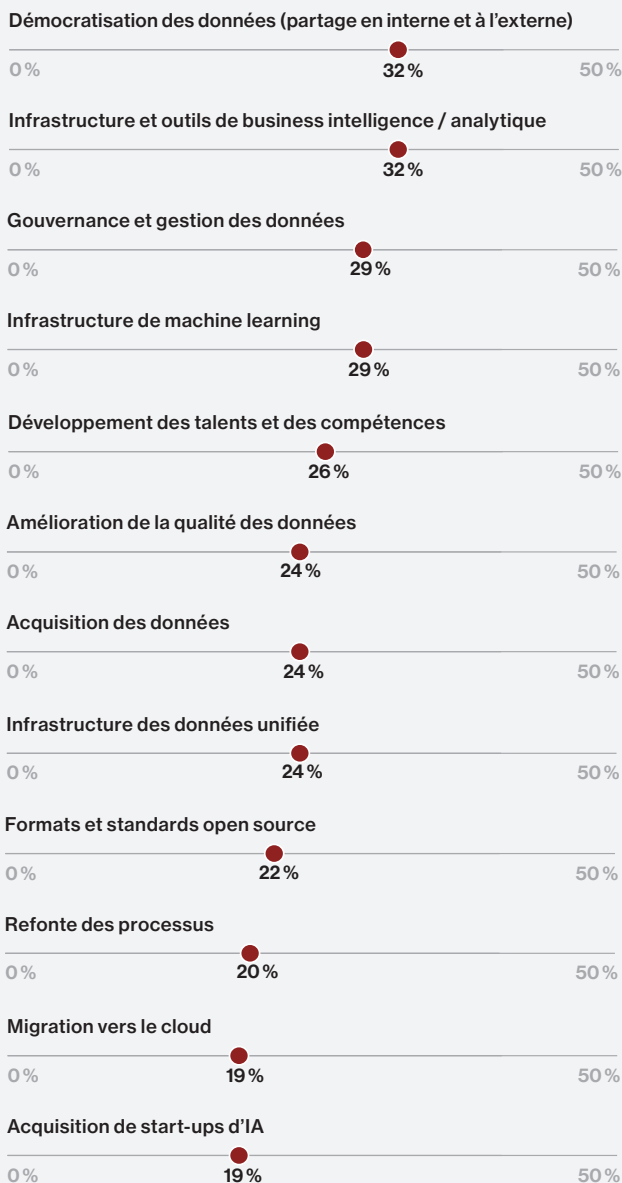
Au cours des dix dernières années, c'est dans la programmation logicielle que l'on trouvait les postes les plus recherchés et les mieux payés. C'est pourtant l'un des domaines où les outils d'IA générative montrent une grande utilité – avec toutes les inquiétudes que cela peut susciter. Les DSI et les responsables techniques interrogés affichent un point de vue plus optimiste. Ils s'attendent en effet à une hausse de la demande en spécialités telles que l'efficacité opérationnelle et l'IA responsable, et pensent que les développeurs vont pouvoir se consacrer à des travaux moins fastidieux et de plus grande valeur. « Ai-je réellement besoin qu'un développeur passe six heures à taper laborieusement du code dans un moteur, puis trois jours à essayer de le déboguer ? » demande Blyton. « Je vois très bien comment on gagnerait en efficacité en s'épargnant ce genre de choses. »

Certains s'attendent à voir les développeurs accompagner la transformation d'idées brutes en résultats matures et exploitables par l'entreprise. « On imagine sans mal, dans un avenir proche, pouvoir utiliser l'anglais pour générer du code, » prédit Blyton. « Dans un an ou deux, mon workflow pourrait consister à confier à des utilisateurs métier une part de l'analyse en langage naturel, puis à incorporer le code obtenu dans quelque chose de plus robuste. » Pour Stoddard, l'IA accélère une tendance déjà présente : « Il vous faut des gens qui comprennent les maths et la science pour travailler avec ces modèles, des personnes qui ont un mode de pensée profondément analytique. Dans mon domaine, les rôles évoluent et concernent de plus en plus l'analytique, l'exploitation des modèles, l'interprétation des écarts et le code. »

Figure 6 : Les domaines d'investissement décisifs pour les données et l'IA

Lorsqu'on leur demande quels investissements ont été décisifs pour générer des bénéfices de l'IA et du ML, les décideurs évoquent avant tout la démocratisation de l'accès aux données et le développement d'une infrastructure de business intelligence intelligente, qui ont déjà produit des résultats.

Pourcentage de participants ayant classé chaque option parmi leurs trois principaux investissements.



Source : Enquête MIT Technology Review Insights, 2022.



« Au cours des cinq à dix prochaines années, nous allons voir à quelle vitesse nous parvenons à nous adapter. Les entreprises qui échoueront, quelle que soit leur taille, sont destinées à disparaître. »

Noriko Rzonca, Chief digital officer, Cosmo Energy Holdings

Pour la grande majorité des DSI, l'IA est bien plus un copilote qu'un rival pour les humains. Zaharia prédit que les LLM permettront aux cadres d'acquiescer eux-mêmes des données de grande qualité pour prendre des décisions, faisant ainsi gagner du temps aux analystes habituellement chargés de répondre à leurs demandes de rapports ad hoc. Mais il ajoute : « Je ne pense pas qu'ils pourront effectuer des tâches extrêmement complexes demandant beaucoup de planification de façon entièrement automatique. Je ne crois pas que le travail de quiconque se limite aux travaux extrêmement simples qu'un modèle de langage peut accomplir. » C'est aussi l'avis de Stoddard : « L'IA générative permet aux créateurs de tous les niveaux de produire du contenu avec leurs propres termes, » explique-t-elle. « Mais je ne crois pas qu'elle pourra jamais remplacer les humains. Elle va plutôt les assister. »

« En interne, nous considérons l'IA/ML comme un appui, un outil qui aide nos équipes et leur permet de consacrer plus de temps à des activités à valeur ajoutée, » ajoute-t-elle. Chez Adobe, il a fallu mettre en œuvre un processus de gestion du changement pour convaincre les équipes. « Quand nous avons lancé le projet, les collaborateurs étaient sceptiques. Mais quand ils ont vu qu'ils pouvaient vraiment gagner du temps pour des tâches plus rentables voire, dans certains cas, rehausser le profil de leur poste, ils y ont adhéré. »

La démocratisation de l'accès aux outils techniques pourrait concerner la société dans son ensemble. « Je pense à toutes les personnes qui ne savent pas programmer ou utiliser un logiciel de conception. Quelles opportunités s'offrent à elles aujourd'hui ? » s'interroge Carbin. « Cela va abaisser le seuil d'accès aux technologies et à la programmation. Aujourd'hui, des logiciels de conception proposent une interface en langage

naturel pour naviguer dans ce qui était auparavant un outil extrêmement complexe, réservé aux meilleurs professionnels. »

L'adoption de l'IA au sein des équipes ne pourra se faire qu'en tenant compte de l'aversion pour le risque et de facteurs culturels comme la peur de l'échec. « Parfois, des personnes qui ont travaillé ou dirigé des organisations au Japon pendant longtemps, et qui se sont habituées à un certain style, ne voient pas la prise de risque d'un aussi bon œil que leurs collègues arrivés plus récemment, qui proposent constamment de nouvelles idées et ont une véritable appétence pour le changement, » souligne Noriko Rzonca, Chief digital officer, Cosmo Energy Holdings. « À l'échelle d'une entreprise, il faut parfois plus de temps pour s'adapter à de nouvelles initiatives et les exploiter pleinement. Mais heureusement, il se trouve que l'amélioration est un véritable point fort des Japonais. Nous allons nous adapter à l'arrivée de l'IA dans nos activités. Et je suis certaine que nous verrons bientôt des évolutions et des améliorations passionnantes dans le fonctionnement des entreprises japonaises. Je suis convaincue que l'IA a un grand potentiel au Japon. »

Pour Rzonca, les organisations n'ont pas d'autre choix que de miser sur les données et l'IA si elles comptent réussir. « Au cours des cinq à dix prochaines années, nous allons voir à quelle vitesse nous parvenons à nous adapter. Les entreprises qui échoueront, quelle que soit leur taille, sont destinées à disparaître. »

06

Risques et responsabilités

L'IA, et en particulier l'IA générative, s'accompagne de défis de gouvernance qui dépassent potentiellement les capacités des cadres actuels en la matière. Lorsqu'elles exploitent des modèles génératifs qui absorbent et régurgitent toutes les données – sensibles ou non – auxquelles ils sont exposés, les organisations doivent adopter une nouvelle approche de la sécurité et de la confidentialité. Les entreprises doivent également gérer des sources de données à la croissance exponentielle, des données produites par des machines ou de provenance douteuse. Tout cela exige une pratique de gouvernance unifiée et cohérente. Les législateurs et les organismes de régulation sont d'ailleurs très conscients des risques posés par l'IA générative. Cela se traduit par des procès, des restrictions et de nouvelles réglementations, comme la loi sur l'IA de l'Union européenne.

Pour toutes ces raisons, les DSI seraient bien imprudents en adoptant des outils d'IA sans gérer les risques : biais, infraction au droit d'auteur, violations de confidentialité, failles de sécurité, etc. « Au sein des services juridiques, financiers, de la confidentialité des données, de l'éthique et de la conformité [de Shell], nous avons pris le temps d'examiner nos politiques et nos cadres pour les adapter à ces nouvelles réalités, » raconte O'Connell.

« Si l'intégralité de votre modèle commercial repose sur votre propriété intellectuelle, sa protection devient absolument centrale. »

Andrew Blyton, Vice-président et DSI, DuPont Water & Protection

La protection de la vie privée est un enjeu majeur. « La technologie est encore jeune, et il faudra beaucoup de grands ensembles de données pour entraîner, valider, vérifier les modèles et réaliser des analyses de dérive, » prévient Schaefer. Au sein du Département des Anciens combattants, « c'est une véritable boîte de Pandore, car nous devons impérativement veiller à ce que les données médicales ne soient pas exposées. Nous avons lourdement investi dans des ressources cloud hautement performantes, régies et sécurisées par des organismes fédéraux. »

La protection du secret commercial et de la propriété intellectuelle est elle aussi source de préoccupations. « Si l'intégralité de votre modèle commercial repose sur votre propriété intellectuelle, la protection devient absolument centrale, » affirme Blyton (DuPont). « D'innombrables acteurs malveillants rêvent de mettre la main sur notre documentation. Nous sommes toujours inquiets de créer de nouveaux vecteurs pouvant entraîner des pertes de propriété intellectuelle. »

La fiabilité est, elle aussi, un enjeu de gouvernance des données. Les LLM créent du contenu inédit en puisant dans des océans d'informations. Mais ils ne savent pas distinguer le vrai du faux. « Quand un modèle croise une information inexacte ou obsolète, il la mémorise tout de même. Cela veut dire qu'il faut faire extrêmement attention à ce que nous injectons, » rappelle Zaharia. « si vous posez deux fois la même question à ChatGPT, vous recevrez deux réponses légèrement différentes. Pour des personnes qui travaillent dans une entreprise dédiée à la science et à l'innovation, c'est assez perturbant, » commente Blyton.

Les modèles doivent être explicables pour gagner la confiance des parties prenantes et apporter la preuve de la valeur commerciale de cette technologie. Cette explicabilité est d'ailleurs déjà une priorité pour la gouvernance de l'IA et les réglementations qui l'entourent, en particulier depuis que les algorithmes jouent un rôle dans des décisions cruciales

comme l'octroi d'un prêt ou l'évaluation du risque de récidive dans le système judiciaire. Le sujet de l'audit algorithmique suscite un intérêt croissant.

« On comprend très bien que l'explicabilité des modèles utilisés dans la santé soit un enjeu majeur, » explique Schaefer. « Nous travaillons sur des cartes de modèles » – une documentation de gouvernance conçue pour fournir des informations normalisées sur l'entraînement d'un modèle, ses points forts et ses faiblesses – « et sur des outils destinés aux registres de modèles et faits pour apporter une certaine explicabilité. Naturellement, le choix de l'algorithme est aussi un facteur essentiel de l'explicabilité d'un modèle. »

Chez Cosmo Energy Holdings, Rzonca cherche l'équilibre entre une gouvernance robuste de l'IA et l'autonomie des acteurs. « Nous élaborons des règles de gouvernance, nous les mettons en place et nous formons les équipes, » explique-t-elle. « En attendant, nous essayons de proposer des outils faciles à prendre en main pour que les résultats soient rapidement tangibles. » Pour atteindre cet équilibre, elle associe la démocratisation de l'accès aux données et à l'IA à une gouvernance centralisée : « Plutôt que de tout faire moi-même, j'essaie de donner de l'autonomie à nos collaborateurs pour qu'ils prennent conscience de leurs capacités individuelles. Et

« L'IA générative élargit de façon exponentielle les possibilités et les promesses de l'IA. Nous pouvons transformer la conversation entre le créateur et la machine. »

Cynthia Stoddard, Vice-présidente senior et DSI, Adobe

Un copilote créatif : l'IA chez Adobe

Les industries de la création adoptent l'IA générative pour offrir de nouvelles capacités fascinantes. Adobe propose aujourd'hui une suite de modèles de conception d'IA générative baptisée Adobe Fireflux. Ses capacités de création prolongent la puissance de Photoshop, Illustrator ou Express et ouvrent une nouvelle ère pour les offres de l'entreprise.

Pour améliorer la vitesse et la productivité d'Adobe Experience Cloud, Adobe a également mis au point Adobe Sensei GenAI, véritable copilote des workflows d'expérience client.

Il améliore la productivité et l'efficacité de l'entreprise dans des domaines comme la création de contenu personnalisé, l'édition et l'expérience conversationnelle.

« L'IA générative décuple les possibilités et les promesses de l'IA, » affirme Cynthia Stoddard, Vice-présidente senior et DSI d'Adobe. « Les outils sont de plus en plus solides, nous pouvons mener de plus en plus de projets, et ainsi transformer la conversation entre le créateur et la machine pour offrir une expérience plus naturelle et intuitive. »

Adobe utilise également l'IA pour optimiser ses propres processus de back-office. En interne, l'entreprise l'utilise pour trier les tickets d'assistance qui peuvent être traités sans intervention humaine, réduire la charge de travail et résoudre plus rapidement les problèmes des clients. Elle a mis l'IA à profit pour cataloguer les innombrables – et parfois redondantes

– solutions logicielles utilisées par ses

différents services, afin de rationaliser les dépenses et simplifier la pile technologique.

Adobe a également développé une plateforme de « guérison spontanée » qui identifie les problèmes techniques réels ou émergents – typiquement le type de problème qui oblige à réveiller un administrateur au milieu de la nuit – et les corrige automatiquement.

C. Stoddard en est convaincue : « pour notre entreprise, l'IA joue un rôle clé dans les expériences que nous offrons à nos clients. Et à l'échelle de l'organisation IT, nous suivons les mêmes principes pour enrichir les capacités de nos processus et faciliter la vie des personnes que nous servons. »

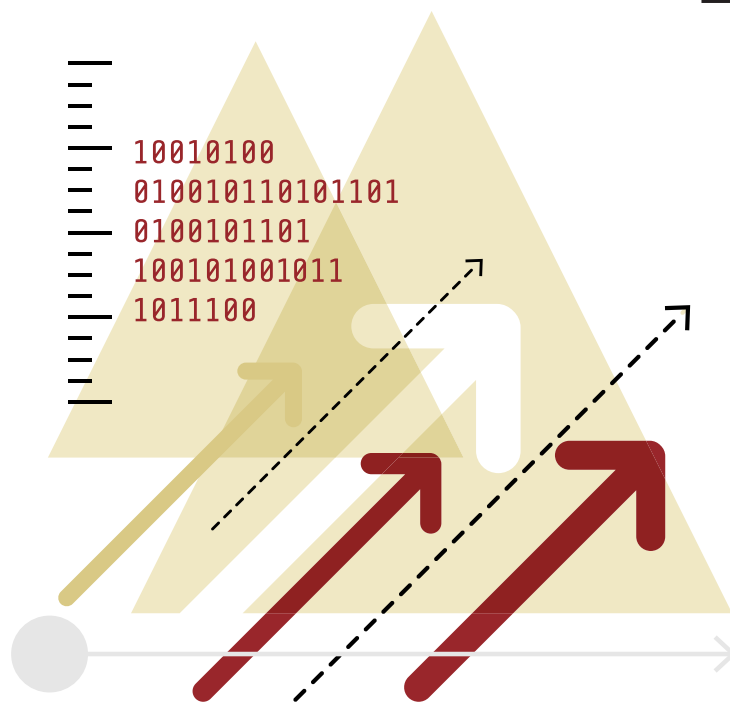
« Nous ressentons le besoin de mettre en place des modèles et des structures de gouvernance très intégrés pour toutes les données et tous les modèles. »

Richard Spencer Schaefer,
Directeur de l'informatique de santé,
Centre médical des Vétérans
de Kansas City

comme mon équipe est autonome, je suis libre de me concentrer sur la sécurité et la gouvernance des données pour éviter tout type de problème dans ce domaine. »

L'importance d'une gouvernance unifiée pour gérer les risques liés à l'IA générative a été régulièrement évoquée au cours de nos entretiens. Au sein du Département des Anciens combattants, « nous ressentons le besoin de mettre en place des modèles et des structures de gouvernance très intégrés pour toutes les données et tous les modèles, » explique Schaefer. « Nous misons de plus en plus sur des outils et des processus très centralisés qui délivrent une solide structure de données. » La gouvernance unifiée a toujours été essentielle, mais l'IA générative en a fait un enjeu plus important encore. « Elle est devenue fondamentale en raison des risques que pose l'introduction de données mal définies et non standardisées dans un modèle, en termes de biais et de dérive, » poursuit-il.

Stoddard souligne l'importance de la diversité des points de vue tout au long du processus de supervision de l'IA chez Adobe. « La diversité dans la supervision est essentielle, et je ne parle pas seulement de la diversité d'ethnicité, de genre ou d'orientation sexuelle. Il faut aussi assurer la diversité des points de vue et des expériences professionnelles des personnes chargées d'évaluer l'impact de l'IA. » Autre sujet clé : la visibilité à l'échelle de l'organisation. Schaefer le dit : « Nous devons mettre en place des outils de gouvernance qui offrent un aperçu visuel des modèles en développement, afin qu'ils puissent être utilisés par la direction et examinés par les parties prenantes à tout moment. »



L'IA constitutionnelle, approche défendue par la start-up Anthropic, impose aux LLM un certain nombre de valeurs et de principes au lieu de compter sur le feedback humain pour orienter la production de contenu.²⁶ Cette démarche guide le modèle afin qu'il applique les normes décrites dans la constitution et s'abstienne, par exemple, de toute réponse toxique ou discriminatoire.

Et malgré les dangers qu'elles comportent pour l'entreprise, les technologies d'IA sont aussi très puissantes pour réduire certains risques commerciaux. « Avec les modèles de langage, il devient plus aisé d'analyser les modèles ou les retours faits par les utilisateurs, » commente Zaharia. « Vous saurez ainsi plus facilement si votre système dévie. » Notre enquête de 2022 a révélé que la sécurité et la gestion des risques constituaient, avec 31 % des réponses, le principal domaine où l'IA avait délivré des bénéfices tangibles. Quant aux impacts positifs attendus pour 2025, il s'agit principalement de la détection des fraudes (27 %), de la cybersécurité (27 %) et de la gestion des risques (26 %).

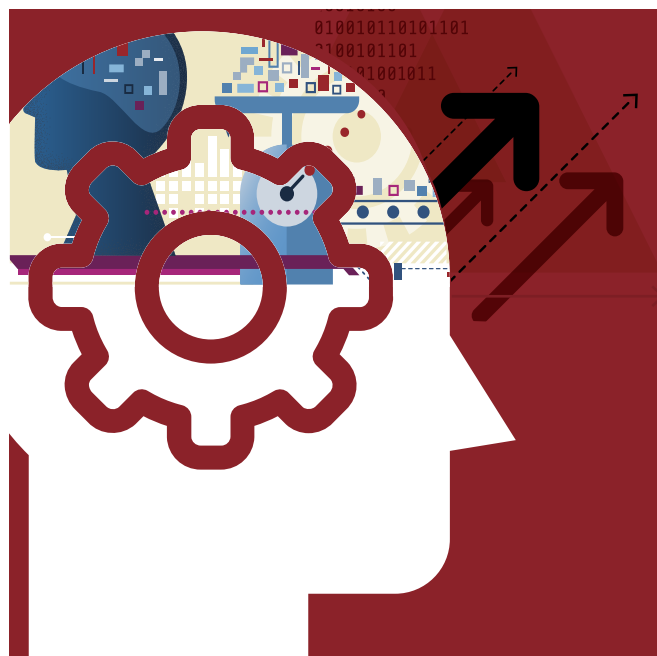
Une nouvelle technologie aussi puissante que l'IA générative s'accompagne de nombreux risques et de lourdes responsabilités. Nos entretiens suggèrent qu'une communauté de professionnels, de start-ups et d'entreprises spécialisées va prendre en main les risques de gouvernance qui la concernent, tout comme les problématiques de développement durable.

07 Conclusion

L'histoire de l'IA est longue et s'étend sur tout le dernier demi-siècle. Jusqu'à récemment, en dehors de projets pilotes ou de contextes très spécifiques comme l'IT et la finance, elle n'avait fait que des incursions limitées dans l'entreprise. L'IA générative est destinée à changer la donne en ouvrant la voie à une adoption véritablement généralisée de l'IA dans l'entreprise.

Rares sont les professionnels aujourd'hui qui considèrent l'IA comme une mode éphémère. Bien souvent, les personnes qui la critiquent sont d'ailleurs aussi les premières à être impressionnées par sa puissance. Mais les décideurs et les experts sont convaincus que les organisations peuvent récolter les fruits de cette innovation tout en gérant ses risques. Les organisations commerciales comme les administrations doivent trouver le fragile équilibre entre l'adoption de l'IA, catalyseur d'innovation et moteur de productivité, et la mise en place de garde-fous visant à atténuer les risques et anticiper les accidents qui ne manqueront pas de se produire.

Tout en gardant ces précautions à l'esprit, les DSI les plus visionnaires entrent avec détermination dans cette ère de l'IA. « Les personnes qui ont connu la révolution de l'ordinateur, puis celle d'Internet, se souviennent nettement de cette époque, » rappelle Blyton. « Si vous faisiez partie de ceux qui savaient se servir d'un ordinateur, votre carrière était assurée. Nous sommes à un tournant similaire : si vous adoptez la technologie, vous en tirerez des bénéfices. »



« Si vous faisiez partie de ceux qui savaient se servir d'un ordinateur, votre carrière était assurée. Nous sommes à un tournant similaire : si vous adoptez la technologie, vous en tirerez des bénéfices. »

Andrew Blyton, Vice-président et DSI, DuPont Water & Protection

À propos de MIT Technology Review Insights

MIT Technology Review Insights est le service de publication sur mesure de *MIT Technology Review*, le magazine sur les technologies le plus ancien du monde. Il est soutenu par la principale institution organisant des événements et publiant des études sur les défis commerciaux et technologiques majeurs de notre époque. Insights mène des études et analyses qualitatives et quantitatives aux États-Unis et en dehors. Il publie aussi des contenus très divers, comme des articles, rapports, infographies, vidéos et podcasts. Et grâce au [Global Insights Panel](#) du MIT Technology Review, qui s'agrandit régulièrement, Insights bénéficie d'un réseau hors pair de dirigeants, d'innovateurs et de leaders visionnaires du monde entier auprès desquels il peut mener des enquêtes et réaliser des interviews.

À propos de Databricks

Databricks est une entreprise axée sur les données et l'intelligence artificielle. Plus de 10 000 entreprises internationales, parmi lesquelles Comcast, Condé Nast et plus de 50 % des entreprises du Fortune 500, s'appuient sur la plateforme Databricks Lakehouse pour unifier leurs données, leurs capacités d'analytique et d'intelligence artificielle. Databricks possède différents bureaux à travers le monde. Son siège social est basé à San Francisco. Fondée par les créateurs d'Apache Spark™, Delta Lake et MLflow, Databricks a pour mission d'aider les équipes en charge des données à répondre aux problèmes les plus complexes. En avril 2023, Databricks a publié Dolly, le premier LLM open source capable de suivre des instructions, finement ajusté à l'aide d'un ensemble de données d'instructions produites par des humains, et utilisable sous licence à des fins commerciales et de recherche. Pour en savoir plus, suivez Databricks sur [Twitter](#), [LinkedIn](#) et [Facebook](#).



Notes de fin

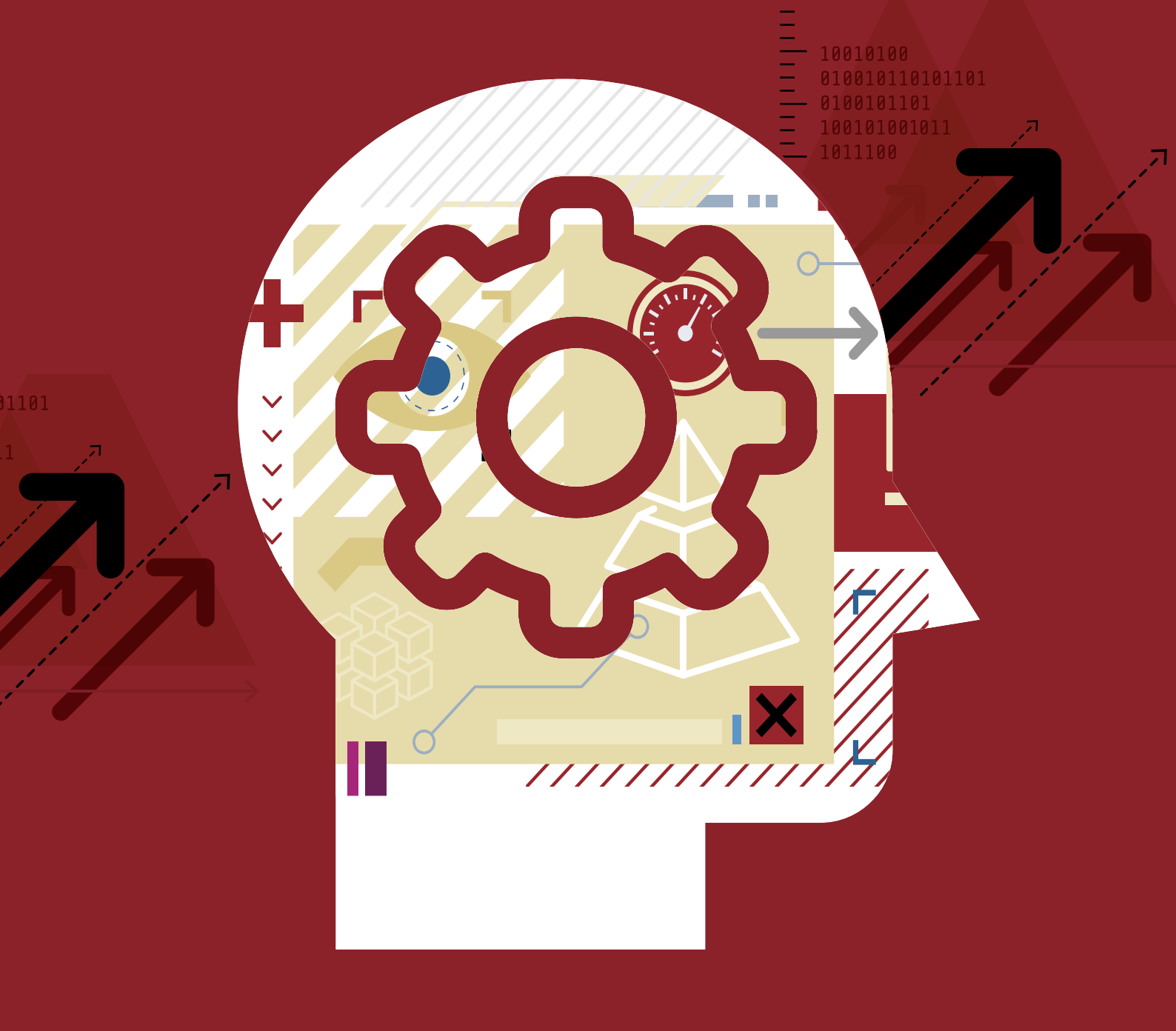
1. « The economic potential of generative AI », McKinsey & Company, 14 juin 2023, <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#/>.
2. « Generative AI could raise global GDP by 7% », Goldman Sachs, 5 avril 2023, <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>.
3. Jim Euchner, « Generative AI », Research-Technology Management, 20 avril 2023, <https://doi.org/10.1080/08956308.2023.2188861>.
4. « CIO vision 2025: Bridging the gap between BI and AI », MIT Technology Review Insights, September 2022, <https://www.technologyreview.com/2022/09/20/1059630/cio-vision-2025-bridging-the-gap-between-bi-and-ai/>.
5. « Retail in the Age of Generative AI », Databricks, 13 avril 2023, <https://www.databricks.com/blog/2023/04/13/retail-age-generative-ai.html>.
6. « The Great Unlock: Large Language Models in Manufacturing », Databricks, 30 mai 2023, <https://www.databricks.com/blog/great-unlock-large-language-models-manufacturing>.
7. « Generative AI Is Everything Everywhere, All at Once », Databricks, 7 juin 2023, <https://www.databricks.com/blog/generative-ai-everything-everywhere-all-at-once>.
8. « Large Language Models in Media & Entertainment », Databricks, 6 juin 2023, <https://www.databricks.com/blog/large-language-models-media-entertainment>.
9. Ibid.
10. Ibid.
11. Ibid.
12. Ibid.
13. Jonathan Vanian et Kif Leswing, « ChatGPT and generative AI are booming, but the costs can be extraordinary », CNBC, 13 mars 2023, <https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html>.
14. Josh Saul et Dina Bass, « Artificial Intelligence Is Booming – So Is Its Carbon Footprint », Bloomberg, 9 mars 2023, <https://www.bloomberg.com/news/articles/2023-03-09/how-much-energy-do-ai-and-chatgpt-use-no-one-knows-for-sure#x4y7vzkg>.
15. Chris Stokel-Walker, « The Generative AI Race Has a Dirty Secret », Wired, 10 février 2023, <https://www.wired.co.uk/article/the-generative-ai-search-race-has-a-dirty-secret>.
16. Vanian et Leswing, « ChatGPT and generative AI are booming, but the costs can be extraordinary », <https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html>.
17. Arthur Spirling, « Why open-source generative AI models are an ethical way forward for science », Nature, 18 avril 2023, <https://www.nature.com/articles/d41586-023-01295-4>.
18. Sharon Goldman, « With a wave of new LLMs, open-source AI is having a moment – and a red-hot debate », VentureBeat, 10 avril 2023, <https://venturebeat.com/ai/with-a-wave-of-new-llms-open-source-ai-is-having-a-moment-and-a-red-hot-debate/>.
19. Thomas H. Davenport et Nitin Mittal, « How Generative AI Is Changing Creative Work », Harvard Business Review, 14 novembre 2022, <https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work>.
20. « Hello Dolly: Democratizing the magic of ChatGPT with open models », Databricks, 24 mars 2023, <https://www.databricks.com/blog/2023/03/24/hello-dolly-democratizing-magic-chatgpt-open-models.html>.
21. « Introducing the World's First Truly Open Instruction-Tuned LLM », Databricks, 12 avril 2023, <https://www.databricks.com/blog/2023/04/12/dolly-first-open-commercially-viable-instruction-tuned-llm>.
22. « Just how good can China get at generative AI? » The Economist, 9 mai 2023, <https://www.economist.com/business/2023/05/09/just-how-good-can-china-get-at-generative-ai>.
23. « Gen AI LLM – A new era of generative AI for everyone », Accenture, 17 avril 2023, <https://www.accenture.com/content/dam/accenture/final/accenture-com/document/Accenture-A-New-Era-of-Generative-AI-for-Everyone.pdf>.
24. « The economic potential of generative AI », McKinsey & Company, 14 juin 2023, <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#/>.
25. « Generative AI could raise global GDP by 7% », Goldman Sachs, 5 avril 2023, <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>.
26. « Claude's Constitution », Anthropic, 9 mai 2023, <https://www.anthropic.com/index/claude-constitution>.

Illustrations

Toutes les illustrations ont été assemblées à partir d'Adobe Stock et des icônes The Noun Project.

Même si nous mettons tout en œuvre pour vérifier leur exactitude et leur fiabilité, MIT Technology Review Insights ne saurait être tenu responsable des informations, opinions ou conclusions exposées dans le présent rapport.

© Copyright MIT Technology Review Insights, 2023. Tous droits réservés.



MIT Technology Review Insights

www.technologyreview.com
insights@technologyreview.com