Intelligence artificielle : 9 bonnes pratiques pour industrialiser les projets

Un grand nombre de projets d'IA échouent faute de cadre méthodologique, de gouvernance de la donnée ou d'implication des métiers. Voici neuf bonnes pratiques pour assurer un passage à l'échelle.

En dépit de l'engouement médiatique autour de l'intelligence artificielle, les entreprises manquent encore de maturité pour mener à bien leurs projets de data science. Beaucoup ne passent pas la phase du POC et, quand c'est le cas, elles constatent souvent un décalage entre les promesses du modèle initial et ses performances en production. Selon le dernier rapport du Capgemini Research Institute, seules 53 % des organisations ont dépassé le stade des projets pilotes. Un chiffre toutefois en hausse puisqu'elles n'étaient que 36 % lors de l'édition de 2017.

La courbe d'apprentissage dans le domaine d'IA reste longue et progressive, freinée par de nombreux obstacles d'ordre technologique, organisationnel ou culturel. Si la volonté est là d'industrialiser leurs projets d'IA pour rendre les livraisons plus systématiques et à un moindre coût, l'approche artisanale, par tâtonnements, reste de mise.

1. Poser un cadre méthodologique

La première erreur serait de considérer un projet d'IA comme un chantier informatique traditionnel. Alors que ce dernier a un début et une fin avec une trajectoire prédictible, un modèle en production doit être supervisé et réentraîné et le projet en lui-même peut être soumis à des changements de cap. « A l'inverse d'un projet traditionnel, les données en IA sont toujours dynamiques, rappelle Nicolas Claudon, chief architect AI & data engineering chez Capgemini.
Le modèle varie dans le temps, en fonction de ces données, ce qu'on appelle la déviation. Sur un modèle de prédiction, cela peut conduire à des écarts de performances. »
« Un projet d'IA est « drivé » par la donnée et peut aboutir à des résultats insoupçonnés, confirme Didier Gaultier, directeur data science & AI chez Business & Decision. Ce qui introduit une forte dose de sérendipité.

Au cours du projet, l'équipe va découvrir des éléments qu'elle ne pouvait anticiper en amont, modifiant son plan de marche. Par ailleurs, des biais peuvent intervenir, ce qui va, là encore, influencer la conduite du projet. » Heureusement, il existe des méthodologies dédiées pour cadrer les projets d'IA.

Développée par IBM dans les années 60, initialement pour le data mining, CRISP va couvrir tout le cycle de vie du projet, de la compréhension du besoin métier au déploiement en passant par la préparation de la donnée, la modélisation et l'évaluation. CRISP est complétée par les approches DataOps pour les travaux d'exploration de données et MLOps pour le déploiement et la maintenance du modèle.

Un projet IA va aussi appliquer les principes DevOps pour favoriser l'intégration et le déploiement

en contenu avec des pipelines génériques de données en fonction des besoins (temps réel, en batch). Il s'agit aussi d'industrialiser packaging et le versioning du modèle et automatiser autant que possible les tests.

2 . Adopter une gouvernance agile

Pour Nicolas Claudon, « la réussite d'un projet d'IA repose sur la capacité à embarquer l'ensemble des parties prenantes : le métier, le département data science, baptisé data lab ou data factory, et la DSI. Cette dernière doit être mis dans la boucle suffisamment tôt pour préparer l'infrastructure, intégrer l'application dans le SI. »

Cette coordination permet, à ses yeux, d'assurer une triple validation à la fois business – le cas d'usage crée-t-il de la valeur ? – data – est-ce que les données sont en nombre suffisant et de qualité ? – et IT – quand et sous quelle forme les livrables seront confiés à la DSI ?

Il s'agit donc de créer des équipes pluridisciplinaires qui portent sur ces enjeux métiers, data et IT. « Il ne doit pas y avoir de rupture au moment du passage de témoin entre le développement du modèle par le data factory et sa mise en production confiée à la DSI », insiste Nicolas Claudon.

Faire travailler de concert les data scientists, l'IT et les métiers n'est toutefois pas chose aisé. Les experts en data science ont souvent pris l'habitude de concevoir leur modèle dans leur coin, en toute autonomie. Par ailleurs, leur formation initiale, avant tout basée sur la modélisation et une approche mathématique, les prépare assez peu au volet IT de l'intégration. On voit ainsi apparaître le métier de machine learning qui va faire le pont entre le data engineer et le data scientist.
Son rôle consiste à optimiser, déployer et maintenir les algorithmes conçus par le data scientist en utilisant les flux de données préparés par le data engineer.

3 . Impliquer les métiers

Autre facteur clé de réussite : l'implication des métiers. « Ils méconnaissent encore les apports réels de l'IA, observe Didier Gaultier. Les métiers ne savent pas précisément ce qu'on peut faire et ne pas faire. Du coup, ils ne sont pas demandeurs. » Il s'agit donc d'acculturer les métiers sur toutes les promesses de l'IA, sans les survendre.

Des séances d'idéation permettent de trouver les cas d'usage pertinents, de l'optimisation de processus existants à la détection des cas de fraude ou des anomalies en passant par l'enrichissement de la connaissance client. Si l'organisation n'est pas encore mature, elle ira sur un cas d'usage simple au ROI rapide, un « quick win » aux gains immédiats. L'entreprise pourra ensuite passer sur des projets au long cours, ambitieux et complexes, où elle se frottera à tous les écueils qui surviendront immanquablement.

Directeur du pôle data de Nexworld, Frederick Miszewski pointe du doigt la gestion du « time to data ». C'est-à-dire le temps qui s'écoule entre l'expression de la demande métier et la validation technique du cas d'usage. Une notion propre aux projets d'IA « Cette période peut créer un effet tunnel. L'équipe de data science revient quelques mois plus tard et le cas d'usage ne correspond plus à celui envisagé par le métier.

En cause : l'absence d'allers-retours entre les deux parties ou une qualité des données insuffisante qui oblige à réviser à la baisse les ambitions. » Pour éviter ce type de déception, il convient, selon lui, d'arriver rapidement à un produit minimum viable ou MVP (minimum viable product) à montrer au client. « Ce qui permet de requalifier demande et de vite rebondir si l'idée de départ s'avère irréalisable ou ne donne pas les résultats escomptés. »

4 . Préparer les données

« La data, c'est le carburant de l'IA », rappelle Didier Gaultier qui établit une corrélation entre la maturité d'une organisation en matière de gouvernance des données et l'aboutissement des projets d'IA. « Avant de parler technologie, il faut questionner la qualité et l'accessibilité des données sachant que le patrimoine data d'une entreprise est souvent siloté et cloisonné. »

Une équipe de data science peut ainsi passer jusqu'à 80 % de son temps à l'extraction et la valorisation des données. « Ce manque de préparation des données retarde au mieux les projets d'IA, au pire elle les fait capoter. »

5. Choisir le bon algorithme

Une fois que les données sont prêtes, quel algorithme choisir ? Il existe une soixantaine des grandes familles d'algorithmes regroupées en six grandes classes. On trouve notamment les algorithmes supervisés. « Ils exigent de disposer de données labellisées, explique Didier Gaultier. Le modèle est entraîné sur des données échantillonnées puis il apprend à les reconnaître par lui-même ». On parle de supervisé profond quand cette approche s'applique au deep learning.

Avec le mode non supervisé, l'algorithme est autonome pour construire sa base d'apprentissage. « Cela suppose un grand volume de données échantillonnées pour qu'il puisse faire des liens et des similitudes ». Dans le mode renforcé, l'algorithme apprend au fil de l'eau mais il a besoin d'un feed-back. « Dans le cas des moteurs de recommandation dans l'e-commerce, ce feed-back est donné par client qui met ou non l'article suggéré dans le panier. »

Comme le montre un tableau comparatif d'un livre blanc de Business & Decision, titré « Intelligence artificielle, restez maître de votre futur !« , le choix du type d'algorithme se fera en fonction du type de données en entrée (structurée, non structurées), du volume de données nécessaire, du mode et de la durée d'apprentissage, de sa complexité, du risque de biais ou des ressources IT nécessaires pour l'exploiter.

La méthode dite de Vapnik, du nom d'un mathématicien russe, permet de trouver l'algorithme idéal. Comme l'a théorisé ce Vladimir Vapnik, la complexité d'un modèle se fait souvent au détriment de sa robustesse, à savoir sa capacité à produire résultats probants dans la durée. Par ailleurs, plus la complexité augmente et plus la traçabilité diminue et l'explicabilité décroit.

6 . S’assurer de l’explicabilité du modèle

Pour éviter l’effet « boîte noire », cette notion d’explicabilité est essentielle. Pour emporter l’adhésion du métier ou de l’utilisateur final mais aussi pour répondre aux exigences réglementaires, il doit être possible d’expliquer comment, à partir de ces données en entrée, un modèle obtient ce résultat en sortie. Une banque doit, par exemple, expliciter les critères d’octroi ou de refus d’un crédit bancaire. « Dans des secteurs sensibles comme la bancassurance ou la défense, cette non-explicabilité du modèle est rédhibitoire, estime Patrick Chable, VP professional services France et Allemagne d’expert.ai.

On peut demander à un être humain pourquoi il agit de telle manière ? Là, un modèle donne un score sans être capable d’expliquer comment il a été obtenu. » Dans le choix de l’algorithme, il s’agit donc trouver un juste équilibre entre performance et transparence. Les arbres de décision ou les régressions linéaires sont aisément explicables, à l’inverse des réseaux neuronaux profonds.
A défaut, il existe différentes méthodes d’interprétabilité qui réduisent cette opacité. Parmi les plus connues, on peut citer LIME (Local Interpretable Model-agnostic Explanations) et SHAP (SHapley Additive exPlanations). Patrick Chable plaide, lui, pour un changement de paradigme.

Aux modèles de machine learning purement statistiques avec un haut niveau d’abstraction, il oppose l’approche symbolique utilisée par expert.ai, spécialiste du traitement du langage naturel. « Pour « comprendre » un document, l’approche symbolique va analyser les relations qui lient différents concepts. On peut ouvrir le capot et voir ce qui détermine un comportement, c’est-à-dire des concepts chaînés entre eux sur le principe du syllogisme. »

7 . S’adosser à l’infrastructure idoine

Se pose ensuite la question du type d’infrastructure qui va héberger le modèle. Pour le traitement des données non structurées, elle repose sur une architecture de type big data qui doit offrir un faible temps d’accès aux données. Didier Gaultier observe que le framework de calcul distribué Spark est de plus en plus utilisé au détriment de Hadoop, « ou bien Hadoop est réduit au stockage et le traitement est assuré avec Spark ». Les bases de données relationnelles restent, elles, pertinentes pour les données structurées.

Autre dilemme : faut-il passer par une infrastructure on-premise ou par un cloud public ? Si une infrastructure en propre peut coûter plus chère que le cloud public, elle offre une portabilité complète, sans risque d’enfermement propriétaire. « Les outils mis à disposition par les hyperscalers ne fonctionnent véritablement que dans leur plateforme », déplore Didier Gaultier. Alternative, une approche hybride consiste à conserver un environnement on-premise tout en rendant possible la bascule dans le cloud.

8 . S’outiller pour automatiser

Les fournisseurs de cloud public comme AWS, Google Cloud et Microsoft Azure proposent à la fois l’infrastructure pour stocker les données mais aussi tout une panoplie de services pour les traiter. Ils commercialisent notamment des solutions comme Google Cloud AutoML ou Azure Automated

ML pour « automatiser » le déploiement de modèles de machine learning en suggérant l'algorithme le plus pertinent à partir d'un jeu de données ou le meilleur pipeline pour un cas d'usage donné.

A côté de l'offre des hyperscalers, il existe des plateformes de data science qui simplifient et accélèrent les déploiements comme Dataiku, KNIME, Viya (SAS), Alteryx, DataRobot ou Domino, référencées par Gartner dans son dernier quadrant magique.

9 . Superviser le modèle

Une fois le modèle en production, le travail ne s'arrête pas là. « Un modèle peut très bien performer puis survient un événement et des données exogènes viennent le pervertir, analyse Frederick Miszewski. Il s'agit de mettre en place un monitoring en continu avec points de mesure pour contrôler ce type de dérives. »

Le modèle doit aussi être ré-entraîné à intervalles réguliers. « Cela suppose d'avoir enregistré les données de production au fil de l'eau, poursuit l'expert. Ce qui n'est pas toujours le cas. » Frederick Miszewski insiste sur l'importance de la gestion de versioning. Les opérationnels doivent savoir quelle version du modèle est déployée mais aussi sur quel jeu de données il a été entrainé.