

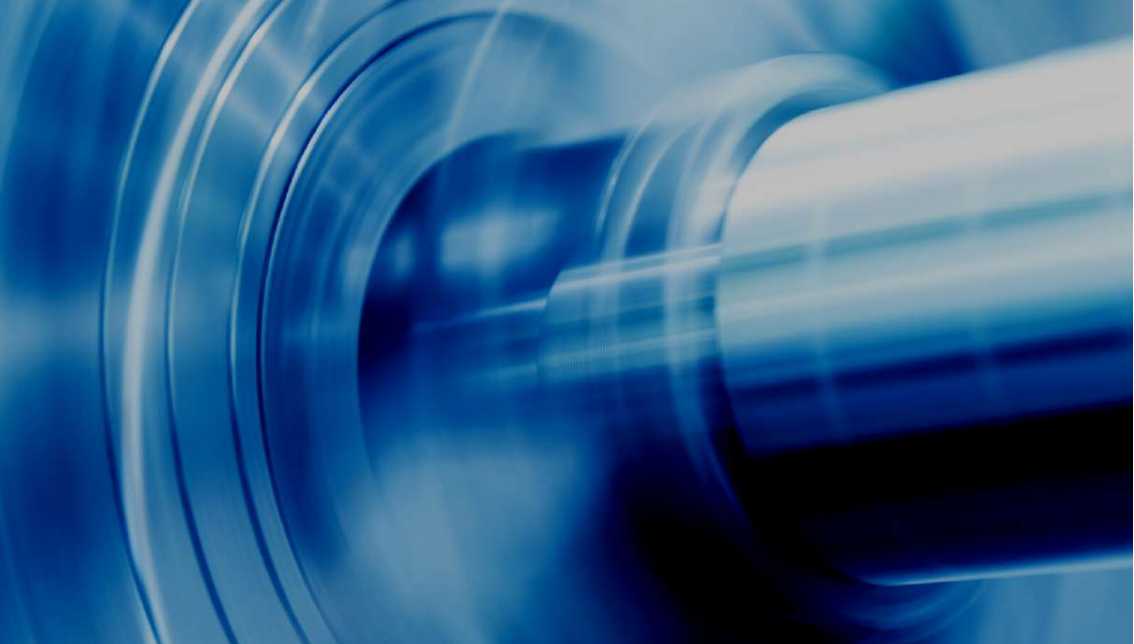


IA

INDUSTRIALISATION ET CAS D'USAGE DANS LES HAUTS-DE-FRANCE

2023





Avec le soutien de :



Direction régionale
de l'économie, de l'emploi,
du travail et des solidarités



Nos partenaires :



Une initiative de :



**PRODUCTION
DU GROUPE DE TRAVAIL
INDUSTRIALISATION
DE L'IA DANS LES
HAUTS-DE-FRANCE
2023**



EDITO

La Cité de l'IA est un réseau d'acteurs régionaux, créée par le Medef Lille Métropole en 2019, pour accompagner les entreprises sur l'Intelligence Artificielle.

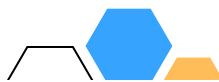
Elle fédère une communauté riche d'hommes et de femmes engagés pour faire de cette technologie un atout pour le développement des Hauts-de-France. Se côtoient ainsi, dans divers formats de rencontres, les mondes économique, académique, scientifique et institutionnel.

Sa singularité est d'être animée et gouvernée par des entreprises utilisatrices de l'IA pour des entreprises utilisatrices, ce qui permet une approche fonctionnelle et pragmatique de ce sujet qui demande encore à être démystifié.

Dans le cadre de ses travaux, la Cité de l'IA recueille les attentes des entreprises et mobilise les acteurs du territoire pour répondre aux besoins exprimés autour de l'Intelligence Artificielle.

Son champ d'intervention est large :

- Sensibilisation du grand public et des entreprises sur les usages de l'IA,
- Accompagnement des entreprises sur l'intégration de solutions. Dans ce cadre, la cité de l'IA est partenaire du programme IA Booster France 2030 afin de déployer le dispositif sur les Hauts-de-France.
- Lobby sur les réflexions réglementaires de la commission Intelligence Artificielle près la Première Ministre, des projets réglementaires de l'UE et du Parlement français,
- Création d'outils pour aider la communauté dans sa compréhension de l'IA,
- Événementiels pour fédérer et favoriser les coopérations.



EDITO

La Cité de l'IA est également à l'origine de la mise en place de plusieurs formations certifiantes sur l'AI portées par des projets cofinancés par l'Etat et la Région.

Le troisième livre blanc proposé par la Cité de l'IA est le fruit des réflexions du groupe de travail "Industrialisation". Il offre aux entreprises la possibilité de s'inspirer de cas d'usages concrets.

Les projets d'IA nécessitent en effet de passer par différentes phases qu'il est important de connaître et d'anticiper pour assurer la viabilité du projet. Toute démarche d'innovation s'accompagne de transformations et de changements pour lesquels il est primordial de poser un cadre. Vous trouverez ainsi dans cet ouvrage les enseignements et bonnes pratiques à connaître, issus de retours d'expériences, pour mener à bien vos projets d'IA.

Nous vous souhaitons une bonne lecture et vous assurons de notre engagement à vous accompagner dans les défis et opportunités que nous ouvre l'Intelligence Artificielle.

Bonne lecture !

Yann Orpin
Président Medef Lille Métropole

Manuel DAVY
Chef de file Cité IA





01. DU POC À L'INDUSTRIALISATION DE L'IA

02. SYNTHÈSE DES RÉSULTATS DE L'ÉTUDE

03. COMMENT APPRIVOISER LA DATA ?

04. A PROPOS D'ALGORITHMES

05. L'INÉVITABLE CHANGEMENT !

06. L'INTÉGRATION TECHNIQUE

07. POUR QUELS RÉSULTATS ?

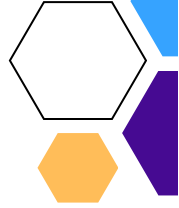
08. ET L'ÉTHIQUE DANS TOUT ÇA ?

09. LE GROUPE DE TRAVAIL



01

DU POC À L'INDUSTRIALISATION DE L'IA



Le premier livre blanc de la cité de l'IA avait abordé la question du processus pour passer de l'idée à la réalisation sur un projet d'Intelligence Artificielle.

Nous avons modélisé le cycle projet en 4 grandes phases : idée, preuve de concept, industrialisation et déploiement. Et pour chacune de ces phases 6 préoccupations clés ont été identifiées : la data, l'algorithme, le résultat attendu, la conduite du changement, l'intégration technique et la question de l'éthique.

Afin de matérialiser chacun de ces sujets, leur impact et de mieux comprendre comment les aborder, nous avons mené une enquête auprès d'une quinzaine de projets réalisés dans des secteurs aussi variés que l'industrie, la distribution, le logement social, les infrastructures, le monde du service,...

Près d'une centaine de questions ont été posées pour chaque cas d'usage afin d'identifier des problématiques communes, comprendre comment elles ont été résolues et essayer de trouver des points communs entre ces projets afin de capitaliser sur leur expérience et de la transmettre dans le cadre de ce livre blanc.

Nous remercions chacun des porteurs de projet d'y avoir consacré sur temps. Certains projets sont confidentiels car concurrentiels, d'autres non et pour ces derniers vous trouverez quelques témoignages dans ce livre blanc.



02

SYNTHÈSE DES RÉSULTATS DE L'ÉTUDE



LES ENTREPRISES FONT AVEC LES DONNÉES À DISPOSITION

Nous n'avons constaté que très peu de construction de données, les approches sont pragmatiques, les entreprises utilisent ce qu'elles ont comme données. En revanche il faut souvent les retravailler, les nettoyer,...

UN USAGE MAJEUR : LA PRÉDICTION

Dans près d'une moitié des cas, l'IA a été utilisée à des fins de prédiction : stock, ventes, fidélisation... avec une amélioration de la performance de l'entreprise qui se traduit financièrement mais qui est parfois difficile à évaluer.

DES APPROCHES PRAGMATIQUES

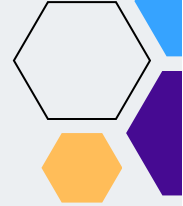
Tous les cas étudiés sont des applications pratiques, très ciblées avec une idée souvent assez précise du résultat escompté ce qui donne des projets relativement rapides de moins d'un an dans la plupart des cas.

MAIS PEU VIENNENT D'ENTRÊPÔTS DE DONNÉES

Rares sont les cas où la donnée est disponible intégralement au sein d'un datalake, dans la plupart des cas il a été nécessaire de mettre en place des tuyaux d'alimentation de données liés à l'usage souhaité.

LE 2e USAGE MAJEUR EST LA PRODUCTIVITÉ

Il s'agit d'un usage évident, celui qui fait le plus craindre pour l'emploi et pourtant dans aucun des cas étudiés il n'y a eu de suppression d'emploi mais plutôt du temps libéré pour d'autres tâches, une meilleure qualité de la production de l'équipe concernée, plus d'attention au client,...



MAIS AVEC PARFOIS UN TEMPS DE RODAGE

Même si cela a pu être rapide dans certains cas, de l'ordre de 10% du temps projet, il a parfois fallu plus de la moitié du temps pour affiner et valider l'algorithme. Et après mise en œuvre, il a fallu dans un cas sur 3, ré-entraîner l'algorithme.

CONDUIRE LE CHANGEMENT EST-CE BIEN UTILE ?

On peut se poser la question tant l'accompagnement dans la plupart des cas a été relativement limité. Peut-être parce que la plupart des chantiers avaient un impact limité sur une fonction précise de l'entreprise et avec un nombre d'acteurs qui l'était aussi.

DES QUESTIONS ÉTHIQUES APPRÉHENDÉES ET EVACUÉES

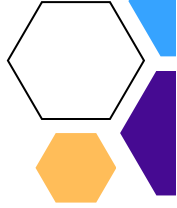
Oui mais en bien, peu de cas ont été confrontés à des questions éthiques, mais dans la totalité des cas concernés, l'équipe projet a appliqué une éthique stricte, anonymisation ou, par exemple, refus d'utiliser la reconnaissance des émotions par une IA,...

En synthèse, les cas d'usage étudiés montrent que l'IA est abordée de façon très pragmatique, sur des sujets précis permettant une meilleure efficacité soit en jouant sur la productivité, soit sur la pertinence. Aucun cas n'est à l'origine d'un nouveau modèle ou d'une stratégie de rupture. Pour beaucoup, ce point de passage permet de mieux appréhender l'usage qui peut-être fait de l'IA. Et s'il y a une recommandation à faire aux lecteurs de cette étude : testez et exercez-vous dans votre entreprise, vous découvrirez peut-être quelques pépites !



03

**COMMENT
APPRIVOISER
LA DATA ?**



Plusieurs critères entrent en ligne de compte pour apprivoiser la data. Identification, confidentialité, labélisation, types de données, fiabilité et accès avant d'atteindre la phase d'industrialisation.

Identification

Il est important de travailler avec les sachants afin d'identifier les données disponibles et ensuite étudier le contenu (exploration des données, quantité, qualité). En effet, toutes les données ne sont pas issues des systèmes d'information, certaines données doivent être récupérées sur d'autres sources. Il est intéressant de noter que seuls 10% des cas étudiés se sont appuyés sur un datalake, la plupart des cas se sont basés sur des mises à disposition de fichiers, des extractions ad'hoc qu'il a fallu industrialiser.

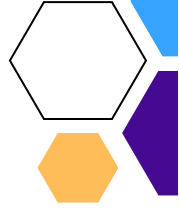
Confidentialité

La notion de sensibilité de la donnée est toujours abordée lors de la sélection, notamment tout ce qui touche aux données à caractère personnel, si besoin une déclaration doit être faite auprès du service approprié (RSSI).

Labélisation

La grande majorité des projets d'IA utilisent des algorithmes dits supervisés nécessitant l'annotation des données pour l'apprentissage. Cela représente un temps non négligeable, et nécessite parfois des profils métiers experts. Lors de notre enquête, 50% des cas étudiés ont nécessité une labélisation des données.





Quels types de données

En majorité, ce sont des données structurées et du texte qu'il faudra analyser. 2/3 des projets se sont appuyés sur des données structurées de l'entreprise, 50% ont utilisé du texte brut. Les données non structurées sont moins fréquentes, seuls 20% ont utilisé de l'image ou du son, mais peuvent apporter une valeur ajoutée importante sur certains cas d'usage.

Fiabilité et données suffisantes

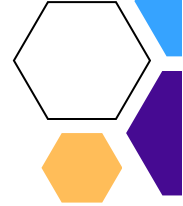
(quantité, balance, biais...)

Une phase d'analyse des données va permettre de déterminer si les données sont significatives et en quantité suffisante pour entraîner un modèle fiable. En fonction des algorithmes, la quantité de données nécessaires varie. Il faut parfois travailler le jeu de données afin de le rendre plus équilibré (balance) et ne pas biaiser l'apprentissage. C'est notamment le cas des données de détection de fraude ou pathologies rares, où les cas positifs sont rares. Et le manque de données peut être pallié dans certains cas, avec la data augmentation ou la génération de données comme l'utilisation d'images de synthèse.

Travail sur les données

(transformation, nettoyage...)

La préparation des données représente souvent le plus gros du travail, au moins 60%, le plus long étant souvent la labellisation. Alimenter le modèle avec des données bien préparées donne de meilleurs résultats. De fait, sur les cas étudiés, les données ont quasiment systématiquement été retravaillées (reconstruites, nettoyées,...).



Accès aux données

Dans la majorité des cas, les données existent sous forme de base de données, de fichiers (via transfert ou partage réseau) ou d'API. Les API peuvent fournir des données de fournisseurs tiers ou des données issues d'objets connectés pour les environnements plus industriels. Pour autant ces données ne sont pas toujours aisément accessibles et 57% des cas étudiés ont été confrontés à des problèmes d'accès aux données.

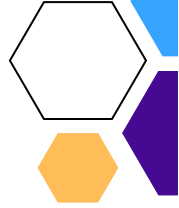
Développement pour récupérer des données supplémentaires

Les techniques de scrapping (extraction des données) permettent de récupérer des données complémentaires disponibles sur des sites web. Le manque de données peut être pallié dans certains cas, par exemple, grâce à des techniques de data augmentation. Ces dernières peuvent notamment rendre possible la création de données supplémentaires en créant des données modifiées à partir des informations existantes (images et son).

Industrialisation de l'accès aux données

Les problèmes d'intégration de données sont récurrents et nécessitent une attention particulière. La généralisation de l'usage des API et des plateformes de données (type Datalake) facilite grandement l'industrialisation. Dans 70% des cas, il a fallu mettre en place des solutions d'alimentation, de transport ou de stockage en phase d'industrialisation.





DataOps

La collaboration des experts data et des experts en automatisation exploitant la richesse de l'outillage existant permet d'avoir une efficacité accrue et durable pour exploiter les données.

Distorsion après le passage en production

Pour éviter ce genre de surprise, il convient d'utiliser des données issues de la production pour l'entraînement, et pas des données travaillées, car la distorsion risque d'être importante. Peu de cas de biais ont cependant été constatés, seulement dans 20% des cas et cela était dû au contexte limité du POC à chaque fois. Il n'y a eu que très peu de cas où en phase de déploiement, une distorsion est apparue sur les données du fait du passage à l'échelle.





“

La Cité de l'IA
Groupe de Travail Industrialisation de l'IA

**COMMENÇONS
DÉJÀ PAR
EXPLOITER LES
DONNÉES A
DISPOSITION**



”



*Jean-Marc Sangnier, Responsable
Service data Promod*

CAS D'USAGE

PROMOD, Premier projet IA, premier ROI

La problématique

La problématique étudiée, chez Promod, est de savoir comment augmenter notre chiffre d'affaires en exploitant les données dont nous disposons déjà ? Nous avons entrepris notre premier projet IA qui est déjà abouti à 90% à l'heure de la rédaction de ce témoignage. Les bénéfices en termes de ROI sont intéressants, mais cela ne s'est pas passé sans difficultés.

Les difficultés

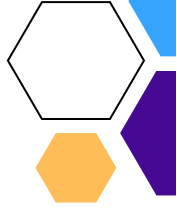
La première difficulté rencontrée est l'approche en rupture avec nos pratiques (mode SaaS, co-construction avec le métier, paiement à l'usage). Nous avons peu de compétences capables de mener à bien ce projet innovant. Par conséquent, nous avons besoin de compétences sur les plateformes data serverless sur le cloud. Enfin, l'une des difficultés, et pas des moindres, fut la conduite de ce changement.

Les bénéfices

Le premier bénéfice a été le transfert de la propriété du projet à Promod, après que la solution ait été mise en œuvre sur le cloud Google par Kantic Analytics. Nous avons bénéficié d'un accompagnement avec transfert de compétences (Data, métier, IT) et d'une capitalisation avec la généralisation de l'architecture pour un Datalake au niveau de l'entreprise. Par ailleurs la démarche de co-construction a permis une intégration naturelle de la solution dans le process de nos équipes marketing (UI/UX).

04

**A PROPOS
D'ALGORITHMES**



Quels résultats attendus ?

(prédiction, clustering, reconnaissance visuelle, traitement du langage naturel...)

La variété des cas d'usage autour des projets d'IA laisse apparaître deux tendances : l'exploitation des données structurées et l'analyse du texte, et la création de nouveaux cas d'usage sur les possibilités offertes par le traitement de la voix, des images et des documents numériques.

Dans notre échantillon, les algorithmes utilisés se sont partagés entre prédiction et modélisation.

Contexte d'exécution de l'algorithme

(Edge computing, objet connecté, cloud, temps réel)

Les typologies d'exécution des algorithmes sont très variées : mode batch, temps réel, fil de l'eau. L'hébergement de l'exécution de l'algorithme est, par contre, plus souvent centralisé (cloud) et plus rarement embarqué (edge).

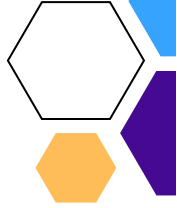
Compétences particulières

La clé de la réussite est une équipe pluridisciplinaire, avec des compétences métier et des profils data science, data et cloud engineer, et des compétences IT comme le développement d'interfaces utilisateur, l'intégration d'exposition des résultats (API par ex.). Dans l'ensemble, les équipes qui ont abordé les projets que nous avons étudiés étaient bien équipées sur le plan de la connaissance des algorithmes, seules 30% ont dû faire appel à des expertises additionnelles.

Quels types d'algorithmes testés ?

En fonction des résultats attendus, on recherche les différents algorithmes possibles pour obtenir les résultats souhaités, puis on teste les candidats en fonction des critères (facilité de mise en œuvre, temps de traitement et puissance nécessaires).





A noter qu'il y a 3 catégories d'algorithmes :

- Des algorithmes sur étagère avec besoin de peu de données mais qui doivent respecter un format attendu.
- La spécialisation d'un algorithme existant (transfert learning) qui sait déjà faire un certain nombre de traitements, auquel on apprend à se spécialiser sur l'objectif attendu. C'est notamment souvent le cas dans le traitement du langage (NLP) et la reconnaissance visuelle, il faut dans ce cas des données labellisées pour l'entraînement d'une toute petite partie du modèle existant. C'est une grande partie des cas que nous avons étudiés.
- La création d'un algorithme spécifique en utilisant des frameworks adaptés, besoin de beaucoup de données annotées, nous n'avons pas rencontré ce cas dans notre étude.

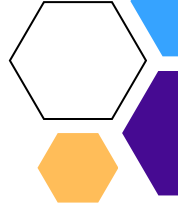
Quels moyens techniques mis à disposition ?

Le besoin d'un environnement disposant d'un outillage riche, et la puissance disponible à la demande font des offres cloud un facilitateur pour les projets d'IA. 90% des cas se sont appuyés sur des moyens techniques dans le cloud chez les 3 grands acteurs du marché à savoir Azure (Microsoft), GCP (Google) et AWS (Amazon).

Quel travail de tuning ?

Tâche fondamentale, le fine tuning de l'algorithme permet de modifier les hyperparamètres du modèle ou de le ré-entraîner afin d'améliorer les résultats générés par le modèle, cela a été fait dans quasiment l'ensemble des cas étudiés (80%).





Temps de mise au point de l'algorithme par rapport au projet

Le temps de mise au point a varié de 10 à 60% du temps du projet dans les cas étudiés, ce qui est non négligeable. Il est recommandé de limiter le temps de mise au point en définissant les critères de résultats attendus tels que le niveau de précision de l'algorithme ou au travers d'un processus formel de validation des résultats.

Difficulté d'industrialisation

L'industrialisation doit être anticipée au démarrage du projet, et en fonction de la cible d'exécution choisie, cette tâche peut représenter une charge plus ou moins importante. Cela peut aller de la consommation d'algorithmes sur étagère (API par exemple) très rapide et facile à mettre en œuvre à la création d'algorithmes spécifiques nécessitant des compétences particulières.

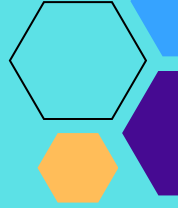
Arbitrage entre pertinence, temps de traitement et adaptation nécessaires pour le passage à l'échelle

En fonction de certaines contraintes, comme le coût, le temps de traitement, la précision du résultat, la puissance disponible, des arbitrages doivent être faits, et si possible anticipés pour éviter un travail important lors du passage en production.

Évolution de l'algorithme après la mise en production

En fonction des cas d'usage, le modèle peut être stable si son écosystème est automatisé et robuste. Certains modèles doivent être réentraînés régulièrement pour mieux prendre en compte les évolutions dans leur environnement de données.



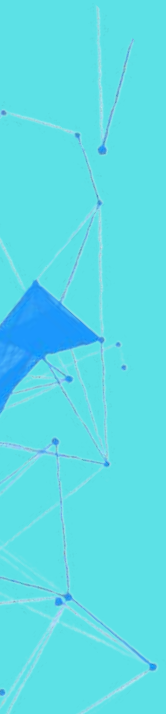


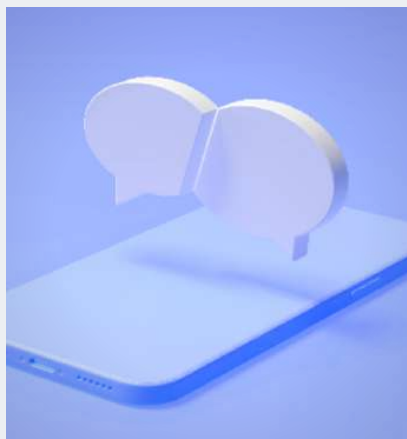
“

La Cité de l'IA
Groupe de Travail Industrialisation de l'IA

ATTENTION AU TEMPS DE MISE AU POINT

”





CAS D'USAGE

SII NORD, **Opinion Viewer** pour résumer les commentaires sur les réseaux sociaux

Opinion Viewer est un projet d'innovation réalisé par le LAB Data & IA de l'agence SII Nord.

Le but de ce projet était de développer une solution capable de classifier et de résumer des commentaires clients issus de réseaux sociaux sur une période donnée. L'objectif final étant de pouvoir générer un résumé de tous les commentaires sur la période et qu'il se rapproche autant que possible de ce qu'aurait fait un humain s'il avait lu tous les commentaires et qu'il en ait fait un résumé.

Afin d'arriver à la génération de ce résumé, une analyse des commentaires a du être réalisée afin de :

- Détecter la polarité générale des commentaires : positifs ou négatifs,
- Identifier les sujets abordés,
- Détecter le sentiment associé à chacun de ces sujets,
- Identifier les mots les plus souvent utilisés dans les commentaires,
- Classifier les commentaires selon une nomenclature produit.

Le travail d'analyse s'est concentré sur l'utilisation et la configuration de différents algorithmes d'IA qui ont ensuite été combinés pour réaliser le pipeline de traitement des données. Les efforts se sont principalement concentrés sur la génération du résumé puisque contrairement à la phase d'analyse, il n'existe pas d'algorithme « sur étagère » permettant de réaliser un résumé humanisé et intelligible à partir de plusieurs textes relativement courts.

Le projet s'est étalé sur deux années et a mobilisé 6 personnes au total : Data Officer, PO Data, Data Scientists et Data Engineers.

05

**L'INÉVITABLE
CHANGEMENT**

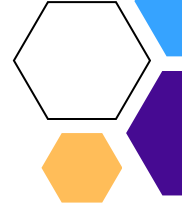


Avant de partager les bonnes pratiques sur la conduite du changement, il est important de comprendre les raisons pour lesquelles certains projets n'ont pas été déployés. Même si cela peut paraître évident, l'absence d'un sponsor et l'absence de gains mesurables représentent les principales difficultés pour embarquer les équipes métiers. Or sur l'échantillon de projets étudiés, l'objectif de départ n'était très clair que dans 2 cas sur 3. Pour les autres il a muté en cours de projet.

Dans l'ensemble ces objectifs étaient soit une question de productivité soit une question de pertinence (de l'offre, de la relation client par exemple,...).

Les indicateurs de réussite, par exemple le gain de temps sur une tâche à faible valeur ajoutée, doivent être mesurés avant le démarrage du projet, et comparés après la mise en place de l'automatisation.

Afin d'amener le changement, des objectifs doivent être définis avec les équipes métiers. Et plus le projet impacte de collaborateurs (dizaines de collaborateurs), plus il a de chance d'être déployé. On pourrait supposer que les projets d'IA aboutissent plus facilement avec des tâches très répétitives plutôt que des tâches très complexes.



Sur les phases de lancement, de pilote et de déploiement, l'implication du sponsor (direction métier, direction financière...) est un facteur clef de succès. Et plus particulièrement lors du pilote où l'engagement des métiers est essentiel avant de déployer.

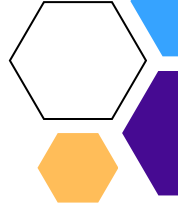
En revanche en phase de ProofOfConcept (POC) où il s'agit de valider la faisabilité et la pertinence de l'algorithme, l'implication des métiers n'est pas forcément nécessaire. Néanmoins cette phase a permis dans un cas sur 2 d'identifier des impacts qui n'avaient pas été imaginés au lancement du projet, la présence des métiers permet alors d'en mesurer plus facilement les conséquences.

En termes de résistance au changement, pour la moitié des cas étudiés, l'impact a été assez transparent pour les équipes métiers, et pour l'autre moitié, l'impact a été positif dans leurs métiers. Ceci peut s'expliquer par la nature de l'échantillon étudié, globalement il s'agissait de sujets très ciblés, les personnes touchées par le projet étaient dans chaque cas en nombre assez limité : des approvisionneurs, des acheteurs, des contrôleurs de gestion, des téléconseillers, des formateurs, des chargés de recrutement. Dans l'ensemble, nos cas d'usage ont concerné essentiellement des métiers intellectuels plutôt que manuels, et peu de cas ont touché les clients de l'entreprise.

Il n'y a eu qu'un cas qui a généré des craintes de remplacement de l'humain par la machine, et un autre pour lequel faire confiance à une boîte noire a été difficile. Dans plus de 80% de cas il n'y a pas eu de difficulté d'appropriation.

Les difficultés d'appropriation ont concerné plus souvent les équipes IT que les équipes métier. La phase d'adoption a généralement été très rapide pour les métiers, en quelques semaines ou quelques mois. En ce qui concerne les équipes IT, des difficultés ont été rencontrées sur les technologies cloud ou les services managés, les faisant passer d'exploitants fournissant des infrastructures à des consommateurs de ressources. Les compétences nécessaires ont plutôt été acquises au travers de formations plutôt qu'en embauchant de nouvelles ressources.

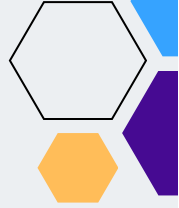




Assez curieusement, moins de 50% des cas ont établi un plan de communication et des actions spécifiques d'accompagnement, alors que plus de la moitié des cas ont fait face à des difficultés d'explicabilité du résultat obtenu, et également plus de 50% ont eu un impact sur la pratique du métier concerné. Dans les cas où la communication interne est intervenue, elle a souvent joué un rôle important pour partager les bénéfices et les performances de l'automatisation.

En synthèse, avec un sponsor impliqué et un suivi des objectifs, on constate davantage une promotion du changement, plutôt qu'une résistance au changement.





“

La Cité de l'IA
Groupe de Travail Industrialisation de l'IA

**LE CHANGEMENT
C'EST CEUX QUI EN
PARLENT LE
MOINS QUI EN
FONT LE PLUS**

”





CAS D'USAGE

Villogia, **Benoit Hemelsdael** **Directeur Stratégie SI** **& Numérique**

Un des chantiers stratégiques de Villogia, Entreprise Sociale pour l'Habitat, avait pour objectif d'optimiser ses points de contacts clients entre 2018 et 2020. Courant 2018, une démarche de design thinking avait été menée avec une quarantaine de collaborateurs pour identifier les leviers numériques à prioriser pour améliorer la satisfaction des clients.

Plusieurs projets ont été lancés fin 2018, par exemple la création d'une application mobile pour faciliter les démarches des locataires. Mais un autre projet plus innovant concernait l'optimisation de notre principal point de contact, le téléphone, en répondant à la problématique suivante : "Comment faciliter la qualification des demandes des clients au téléphone ?" Et si l'Intelligence Artificielle pouvait nous aider ? En effet, plus de 200 motifs permettent de qualifier la demande d'un client.

Début 2019, un hackathon a été réalisé dans les locaux de Vilogia afin de sélectionner les meilleurs partenaires potentiels pour nous accompagner sur ce projet. Le jury était composé des membres du comité de direction, et les démonstrations ont eu lieu dans le hall du siège et étaient ouvertes aux collaborateurs. En 48h, les experts de l'IA avaient su impressionner les spectateurs, dont des membres de notre centre de relation clients, et la faisabilité du projet était confirmée.

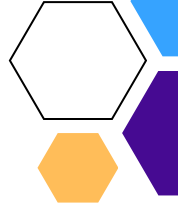
Durant cette période, une sensibilisation sur l'IA a été organisée auprès des managers du Groupe, quels usages dans notre secteur ? Quels impacts dans le service client ? Quels apports potentiels ?

La plus grande crainte sur le changement opéré concernait nos clients, quelles allaient être leurs réactions ? Un benchmark avait donc été mené sur l'usage des serveurs vocaux intelligents (CallBot) dans d'autres secteurs d'activité et dans l'immobilier. Aussi pour faciliter l'adhésion de la direction métier, un déploiement progressif a été opéré, des indicateurs (KPI) ont été développés afin de mesurer l'efficacité et l'impact du callbot sur le temps d'attente, le taux de décroché, le taux d'automatisation de qualification des demandes... En parallèle des API ont été développés pour ne pas poser des questions inutiles au client quand on a la data dans nos SI.

Enfin et surtout, la solution développée est désormais paramétrée directement par le centre de relation client, les attentes des clients évoluent, il faut de la réactivité et de l'instantanéité, l'IA y contribue fortement.

06

**L'INTÉGRATION
TECHNIQUE**



L'intégration doit être bien prise en compte et bien préparée dès le lancement du projet. En effet, elle peut impacter l'architecture des SI, les compétences internes ou externes à recruter, et les coûts d'infrastructure et d'exploitation.

La majorité des projets (57%) ont réalisé une analyse d'impact pour évaluer la capacité et les difficultés d'intégration dès le démarrage du projet. Et dans l'ensemble, les équipes techniques chargées de l'intégration ont été impliquées très tôt dans le processus projet.

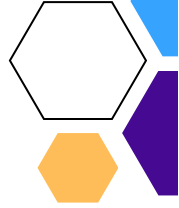
Les solutions intégrées ou semi-intégrées dès la phase de ProofOfConcept (POC) ont plus de chance d'aboutir et d'être déployées que les solutions totalement isolées. Néanmoins, 50% des cas étudiés ont été réalisés de façon isolée.

La phase de POC permet surtout de mener beaucoup d'apprentissage, de tester les algorithmes, d'évaluer la pertinence des modèles, ainsi que la potentialité des technologies.

La gestion du volume des données est rarement problématique. Dans 80% des cas cela n'a posé aucun problème, sauf quand ce volume est insuffisant pour apporter des résultats probants sur certains cas, il convient alors d'être un peu plus patient ou de revoir sa méthodologie de captation des données.

Le canal de captation est à prendre en compte dans l'intégration. Par exemple, un agent conversationnel de type chatbot sur un site web ne peut être pris en compte comme un callbot couplé à un système de téléphonie. En termes de flux, quand les canaux (ex: audios et vidéos) nécessitent des échanges temps réels, les API sont recommandés, mais pour les autres cas d'usage les traitements en temps différé peuvent suffire.





En revanche, dans 1 cas sur 2, il a fallu modifier les systèmes existants pour intégrer correctement la solution. Cela s'est fait de façon fluide, mais a tout de même ralenti les projets dans un cas sur 3, plutôt pour des questions de disponibilité de ressource ou d'organisation complexe que pour des questions purement techniques.

Les impacts IT (techno Cloud, service managés, API) peuvent entraîner des dispositifs spécifiques en termes d'exploitation, cela a été le cas dans 1 projet sur 3.

Peu de difficultés de stabilité ou de performance des solutions ont été rencontrées, il y a juste eu parfois des problématiques de sécurité d'accès aux données à résoudre, et dans la plupart des cas, les solutions déployées semblent robustes aujourd'hui, elles reposent souvent sur des composants réutilisables et scalables.

La majorité des entreprises ont constaté des coûts d'exploitation et d'infrastructures conformes ou inférieurs aux coûts budgétés.

En conclusion une bonne anticipation permet de résoudre assez aisément les points techniques sur des chantiers ciblés. Sur des chantiers plus vastes que ceux étudiés, cela serait probablement différent.





*Olivier Burel, Responsable Innovation
Voies Navigables de France*

CAS D'USAGE

VNF, Projet Fluviate Quai

Nous nous modernisons grâce aux données, à l'IA et aux objets connectés

La problématique à laquelle nous étions confrontés était d'optimiser le trafic fluvial en remontant des informations sur l'occupation des quais de stationnement.

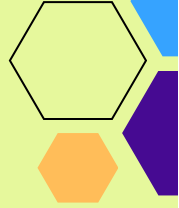
Pour faire de cette idée une réalité, nous avons fait face à plusieurs difficultés. La première, dans notre organisation, consistait en un changement d'approche avec la nécessité de collaborer entre l'informatique classique de gestion, l'informatique industrielle (c'est-à-dire les automatismes présents sur les équipements de la voie d'eau) et le génie civil.

Puis, dans ce projet multiforme, nous nous sommes heurtés à la lourdeur de certains processus administratifs (par exemple pour l'achat de matériel), à l'indisponibilité des données pour entraîner le modèle, ou encore à des questions de propriété intellectuelle des algorithmes et d'engagement de résultat.

Néanmoins nous avons beaucoup appris, tout d'abord sur la nécessité de disposer de plateformes de données communes, mais aussi sur la collaboration entre informatique de gestion et informatique industrielle autour des objets connectés. L'accompagnement de Kantic Analytics nous a aidé dans la vulgarisation de l'usage des technologies d'IA dans nos projets.

Enfin ce premier projet a abouti à la mise en place d'un démonstrateur d'identification des convois et a ouvert des perspectives pour de nombreux autres cas d'usage.

Le test est validé !



“

La Cité de l'IA
Groupe de Travail Industrialisation de l'IA

**LES
SOLUTIONS
INTÉGRÉES DÈS
LA PHASE DE
POC ONT PLUS
DE CHANCE
D'ABOUTIR**

”



07

**POUR QUELS
RÉSULTATS ?**



Le résultat attendu au travers des cas d'usage étudiés était très équilibré entre ceux qui visaient de la productivité, délai de traitement d'une demande client, temps d'attente au téléphone, analyse des avis client, production de documents... et ceux qui visaient une meilleure prédiction, tels que les prévisions de vente, l'optimisation du stock, le risque de départ d'un client, ou l'ajustement des moyens à la prévision d'activité.

Les cas d'usage sont variés et les objectifs étaient très précis dans 2 cas sur 3, néanmoins dans le tiers restant, l'objectif a pu muter en cours de projet au regard de la réalisation du POC.

Pour la plupart néanmoins, personne n'avait une idée précise des gains réalisables, seuls 1 cas sur 5 avait une idée claire et chiffrée des bénéfices escomptés.

Dans la plupart des cas le chantier a été engagé sur la base d'un sponsorship fort du bénéficiaire mais aussi, à un degré moindre, car le coût de test était relativement faible.

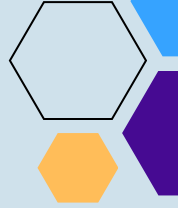
Très rapidement pour la plupart des projets, le POC a pu offrir une précision de 70 à 80% de l'algorithme, permettant de valider le cas d'usage.

Peu de projets dans notre échantillon n'ont pas donné le résultat escompté, ce qui fait que 80% des cas étudiés sont en production et celle-ci a systématiquement confirmé les enseignements du POC. Ce taux de réussite est néanmoins à prendre avec précaution du fait du nombre limité de cas étudiés, notre étude ne se veut pas exhaustive. Nous nous sommes concentrés sur des projets ayant abouti.

Pendant un tiers des cas ont eu des difficultés pour mesurer le résultat, dans un cas sur 2, il a fallu mettre en place une solution pour mesurer le résultat.

Suite au déploiement de la solution, tous assurent un contrôle régulier de la pertinence de l'algorithme et du résultat obtenu.





“

La Cité de l'IA
Groupe de Travail Industrialisation de l'IA

**QUAND LE BUT
EST CLAIR ET
LE RESULTAT
PROUVÉ,
BEAUCOUP DE
DIFFICULTÉS
S'ESTOMPENT**

”





CAS D'USAGE

NOLEJ AI, le nouvel outil pédagogique pour les enseignants

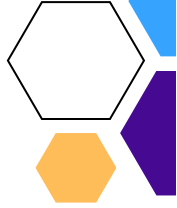
NOLEJ est une startup EdTech des Hauts de France spécialisée dans le traitement du langage naturel avant l'émergence des IA génératives telles que ChatGPT. Elle commercialise aujourd'hui un produit mature NOLEJ AI, co-élaboré avec des milliers d'enseignants dans une démarche d'intelligence collective.

L'outil NOLEJ AI permet de créer davantage d'engagement en transformant en modules interactifs tous supports pédagogiques statiques existants tels que du texte, de l'audio et de la vidéo. Véritable gain de temps, ses modules interactifs sont créés jusqu'à 50 fois plus rapidement qu'avec les méthodes traditionnelles. Fort de 5 années d'expertise en intelligence artificielle, en pédagogie, en psychologie cognitive et en intelligence collective, l'équipe de NOLEJ s'est spécialisée dans la fiabilité des données générées par l'IA. En créant douze activités interactives alignées avec le contenu du document source, NOLEJ AI évite les biais, les informations trompeuses et les dérives des IA génératives. Pour plus de praticité, NOLEJ AI vise à épouser l'écosystème numérique existant de millions d'enseignants avec des intégrations directes dans les principaux systèmes de gestion de l'apprentissage du marché tels que Moodle et Canvas, ainsi que dans les classes virtuelles de Google Classroom et de Microsoft Teams. L'outil fonctionne déjà dans plusieurs langues, dont l'anglais, le français, l'espagnol, l'allemand, l'italien, le néerlandais et le portugais.

NOLEJ AI est utilisé par 59 000 enseignants dans plus de 800 établissements et 10 pays. L'entreprise a signé un accord avec la Région Île-de-France pour fournir son outil d'intelligence artificielle éthique à tous les enseignants de ses 625 lycées. Face à 3200 start-ups Deeptech et à 900 startups EdTech de l'ASU-GSV Cup aux Etats-Unis, elle a été classée dans le top 20. Ce qui lui a valu d'être distinguée par le Label d'Excellence du Conseil européen de l'innovation.

08

**ET L'ÉTHIQUE
DANS
TOUT ÇA ?**



Parmi les cas étudiés, et du fait de leur champ d'application, peu ont été concernés par des questions éthiques. Et ceci pour plusieurs raisons : pas de risque physique sur les biens ou les personnes dans les cas étudiés, très peu de données sensibles manipulées,... Nous étions sur des cas d'usage de gestion d'entreprise somme toute assez standards.

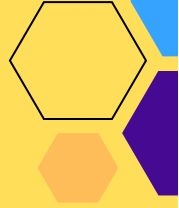
Néanmoins, il y en a eu quelques-uns et il est intéressant de voir la manière dont cela a été traité, on peut parler d'évitement éthique. Le choix a clairement été fait de refuser toute approche via l'IA contraire à l'éthique.

A titre d'exemple, un cas d'usage est basé sur la voix, il a clairement été exclu d'utiliser une IA qui pourrait déterminer les émotions à partir des intonations de la voix. Non pas car cela reste assez aléatoire à date, mais pour des questions de principes éthiques. De même, dans les cas où des données personnelles sont utilisées, elles sont systématiquement anonymisées.

Nous avons voulu savoir si des référentiels publics sur l'éthique et l'IA ont été utilisés. C'est le cas de 2 équipes, l'une s'est appuyée sur le référentiel du Syntec, l'autre sur le Stanford Human AI. Pour partie, cela est dû à la non nécessité de les employer dans le cadre du projet considéré mais aussi parce qu'ils ne sont pas connus.

Il s'agit probablement là d'un point de progrès important dans le cadre de la mise en place des projets d'IA. Même si la sensibilisation existe, la prise en considération stricte des principes éthiques est loin des préoccupations des équipes chargées de la mise en place. Ceci peut s'expliquer par la phase dans laquelle nous nous trouvons sur l'usage de l'IA, même si des projets existent dans de nombreuses entreprises, nous en sommes à une étape qui est encore exploratoire.





“

La Cité de l'IA
Groupe de Travail Industrialisation de l'IA

**GLOBALEMENT,
BIEN INTEGRÉE,
L'ÉTHIQUE
RESTE A
APPRÉHENDER
DE FAÇON PLUS
FORMELLE**

”





CAS D'USAGE

SII NORD, **CulturIA, pour sélectionner** **le bon candidat**

CulturIA est un projet d'innovation réalisé par le LAB Data & IA de l'agence SII Nord.

Le cas d'usage de ce projet est l'aide à la décision lors de la sélection de candidats en recrutement. L'idée directrice est la capacité, grâce à l'IA, de connaître la culture d'une entreprise et celle des candidats afin de calculer un score de correspondance donnant une indication sur la capacité du candidat à s'intégrer dans l'entreprise.

Il a fallu avant tout être en capacité de faire ressortir une personnalité, une culture d'un CV. La complexité de cette tâche réside dans la diversité des éléments décrivant un CV : une photo, une potentielle phrase d'accroche, des icônes, des logos, du texte, des jauges de niveau de compétences, un certain agencement, des couleurs caractéristiques, etc.

Durant tout le processus d'extraction de l'information, il était important d'éviter au maximum que des biais importants s'immiscent dans nos données. Au-delà des biais issus des jeux de données, une grosse partie du projet a été consacrée à la partie éthique car il est évident que ce genre d'outils peut facilement être générateur de problèmes.

Différents scoring utilisant des matrices publiques ont donc été réalisés et ont permis de mettre en évidence les points à prendre en compte dès le traitement des données. Cette vigilance a été nécessaire tout au long du projet afin que le scoring établi par l'outil ne contienne aucune discrimination.

Le projet s'est étalé sur deux années et a mobilisé 5 personnes au total : Data Officer, Responsable Recrutement agissant en tant que PO, Data Scientists et Data Engineers.

09

LE GROUPE DE TRAVAIL INDUSTRIALISATION ET CAS D'USAGE DE L'IA



LE GROUPE DE TRAVAIL INDUSTRIALISATION ET CAS D'USAGE

Remerciements aux membres du groupe de travail ayant contribué à l'élaboration de ce livre blanc.

Vincent BEHAGUE, ADIX

Mathieu DESMAREZ, ASCENT WAI

Benoit HEMELSDAEL, VILOGIA

Stéphane LECOEUICHE, IMT NORD EUROPE

Guillaume LEPRINCE, GROUPE SII

Justine NOSLIER, GROUPE SII

Michel PHILIPPART, EDHEC

Véronique PROUVOST, ARC INTERNATIONAL

Jacky RENNO, KANTIC ANALYTICS

Jean-Michel ROUSSEL, PAYBACK GROUP

Martin SINGER, EY

Vincent SIX, JUNIA

Pilote du groupe de travail

Vincent BEHAGUE, ADIX



LA CITÉ DE L'IA

CAMPUS ENTREPRISES ET CITÉS
40 RUE EUGÈNE JACQUET
59700 MARCO-EN-BAROEUL
IAHDF.ORG

RÉALISATION

KAYAK COMMUNICATION

IMPRESSION

LA REPRO DE MARCO

IA

INDUSTRIALISATION ET CAS D'USAGE DANS LES HAUTS-DE-FRANCE

2023



LA CITÉ DE L'IA
CAMPUS ENTREPRISES ET CITÉS
40 RUE EUGÈNE JACQUET
59700 MARCO-EN-BAROEUL
IAHDF.ORG