FICHES PRATIQUES
IA RESPONSABLE

IMPACT AI

FICHES PRATIQUES IA RESPONSABLE

La culture de la responsabilité se développe dans les entreprises en même temps que la croissance du développement de l'IA pour répondre à des exigences sociétales fortes.

Pour autant le chemin est encore long et les difficultés multiples, ce qui suppose une approche spécifique pour aider les différentes parties prenantes à naviguer de manière responsable dans le développement de systèmes d'IA tout au long de leur cycle de vie.

Les personnes qui se concentrent sur le développement et la livraison de tels systèmes doivent veiller à prendre des mesures suffisantes pour étendre l'IA responsable a l'ensemble de leur organisation.

En définitive, l'IA responsable consiste à provoquer un changement culturel au sein d'une organisation.

La mise en œuvre d'une IA responsable nécessite des changements importants au niveau du leadership, de la gouvernance, des processus et des talents.

Ce livret est l'aboutissement des travaux réalisés par les partenaires d'Impact AI lors du cycle 2021 du Groupe de travail IA Responsable.

Nous avons voulu capitaliser sur nos échanges et regards croisés pour explorer et retranscrire par étape du cycle de vie d'un système d'IA les enjeux, difficultés et bonnes pratiques à destination des data scientists, des ingénieurs Machine Learning, des développeurs, des designers, etc... tout en reconnaissant que le développement des systèmes d'IA passe souvent par certaines de ces phases de manière itérative.

L'objectif de ces fiches est d'illustrer comment évaluer la valeur, les risques et les impacts d'un projet d'IA dès la phase d'idéation jusqu'aux usages.

Chaque fiche représente une étape particulière du cycle de vie de l'IA en se concentrant sur les points suivants :

- *Processus – Outils – Pilotage*
- *Acteurs concernés*
- *Points de vigilance*
- *Bonnes pratiques identifiées*

Cycle de vie de l'Intelligence Artificielle

Ces fiches pratiques constituent un moyen formel de **stimuler les conversations et la réflexion nécessaires pour anticiper et atténuer les risques des systèmes d'IA.**

Elles se concentrent sur le développement de ces systèmes dès leur conception, depuis les processus d'idéation jusqu'à la mise en production et les usages associés.

Nous sommes conscients que chaque cas d'utilisation présente un contexte et un ensemble de défis uniques et que ces préconisations devront tenir compte de ce contexte.

En outre, les organisations doivent envisager le développement de systèmes d'IA dans le contexte d'un environnement réglementaire en évolution rapide. Les présentes fiches et leurs lignes directrices ne doivent donc pas être considérées comme un outil permettant d'atteindre la conformité réglementaire ou juridique.

Les équipes impliquées doivent toujours travailler avec les services internes concernés pour s'assurer que les systèmes d'IA qu'elles créent, respectent toutes les lois et réglementations applicables dans les juridictions où le système sera développé, utilisé ou commercialisé.

Nous pensons que ces fiches pratiques aideront les équipes à créer une synergie plus profonde entre les performances, les objectifs organisationnels et les valeurs. Tout au long du cycle de vie du développement d'un système d'IA, les parties prenantes doivent garder à l'esprit les valeurs de leur organisation et les principes éthiques de l'IA, dans la mesure où leur organisation les a formalisé. Inévitablement, des problèmes surviendront qui obligeront les équipes à faire preuve de jugement.

Dans ces situations, les équipes doivent utiliser les valeurs et les principes de leur organisation comme une étoile polaire pour guider leur réflexion. Elles doivent également rechercher des perspectives diverses à l'intérieur et à l'extérieur de leur organisation afin d'obtenir l'avis de toutes les parties prenantes concernées.

Bonne lecture,

Le collectif Impact AI, groupe de travail IA Responsable

PHASE 1

IMPACT AI

IDÉATION ET DÉFINITION DE CAS D'USAGE IA

OBJECTIF

Faire émerger et définir les cas d'usage en intégrant la dimension sociétale, l'impact sur les personnes et l'environnement, les effets de bord possible

PROCESSUS - OUTILS - PILOTAGE

Il n'y a pas de processus type car cette phase peut être très **divergente** mais il pourra y avoir :

- Une activité de **veille** avec des sources d'informations diverses (innovation technique, science fiction, besoins sociétaux et environnementaux, pièges ou ratés passés en interne et en externe) en particulier sur les questions éthiques
- Des phases de brainstorm entre acteurs et du design thinking
- Des phases de **prototypage** et de démonstration pour montrer la valeur et faire mûrir les idées
- Des tests utilisateurs en intégrant les parties prenantes concernées
- La définition du processus de qualification/priorisation du cas d'usage.

POINTS DE VIGILANCE

Cette phase **itérative** n'aboutit pas toujours à un succès et il est parfois difficile de détecter suffisamment tôt et de **capitaliser sur les échecs.**

Cela entraîne une difficulté à itérer rapidement (savoir arrêter suffisamment tôt, rebondir et repartir sur une autre voix).

Etant donné la diversité des acteurs, la divergence de point de vue ou de compréhension technique,la communication, le partage des idées n'est pas toujours facile.

Il est important d'anticiper le passage à l'échelle et la réplicabilité des cas d'usage, ce qui est difficile lorsque l'on s'adresse à des entités de culture ou de maturité variables en termes de technologie, de réglementation, d'anticipation et d'accompagnement du changement.

ACTEURS CONCERNÉS

Cette phase nécessite une équipe **pluridisciplinaire** dotée de l'éventail de compétences nécessaires à l'idéation.

Les acteurs internes de l'entreprise (équipes métier, recherche, innovation, informatique, au niveau du groupe ou d'une filiale) peuvent venir de différents horizons (business, technique, juridique, sociologique, ressources humaines, communication).

Mais ils peuvent également être des acteurs **externes** : clients, partenaires, utilisateurs, citoyens, régulateurs, politiques, artistes.

Ces acteurs sont en général pilotés par une équipe data science qui centralise les idées et demandes.

CONCEPTS CLÉS

Besoin Métier | Valeurs éthiques | Parties prenantes diverses

LES BONNES PRATIQUES IDENTIFIÉES

- Définir les **rôles et responsabilités** assez vite pour la sélection des cas d'usage avec des critères
- Adopter un point de vue le plus général et **ouvert** possible dans la réflexion dès le départ en faisant se croiser des acteurs divers
- Cultiver, former, sensibiliser les salariés - au-delà des équipes d'IA – aux avantages attendus, conditions de fonctionnement et risques associés à l'utilisation de techniques d'IA.
- Traduire les principes éthiques en pratique, par exemple avec des questions concrètes dès l'idéation.

"Pour partir d'un bon pied, il convient de s'interroger, dès la définition du problème que l'on souhaite résoudre, sur les impacts positifs et contreparties négatives éventuelles en termes de valeur business, d'organisation, d'implication de l'utilisateur et d'interprétabilité des résultats, afin d'en tenir compte dès les premières phases de prototypage et de démonstration aux parties concernées. "

Claude le Pape - Intelligence, Optimization & Analytics Fellow, IoT & Digital Offers – Schneider Electric

PHASE 2

IMPACT AI

QUALIFICATION DE CAS D'USAGE IA

OBJECTIF

Qualifier les cas d'usages en intégrant la dimension responsable, pour passer d'une idée à un projet concret à lancer.

- Évaluer le **bien-fondé** du développement du cas d'usage IA en tenant compte des valeurs organisationnelles et des objectifs commerciaux
- Évaluer l'**impact** potentiel du produit avec l'aide d'experts du domaine et des groupes potentiellement concernés
- Constituer une **équipe** reflétant diverses perspectives avec des rôles et des responsabilités clairement définis pour le cycle de vie du produit d'IA.

PROCESSUS - OUTILS - PILOTAGE

La qualification pour intégrer la dimension responsable doit intégrer une phase de **questionnement** complémentaire (par exemple basée sur des principes éthiques) et une organisation associée (par exemple avec un comité de gouvernance qui valide les étapes clés). Les documents de référence, tels qu'une **Charte** Éthique IA ou IA Responsable, ont pour but d'assurer l'alignement de principes et valeurs sur l'ensemble du cycle de vie d'un projet d'IA. Les équipes pilotant les projets d'IA se réfèrent à ce document afin d'assurer la mise en place des principes et valeurs consensuelles au sein de l'organisation. Par exemple, les critères de fiabilité et d'équité des modèles.

La **formation** des acteurs doit suivre la mise en place de la démarche et la conduite du changement. De nouveaux réflexes peuvent alors se développer : challenger les choix d'algorithmes et l'objectif commun, assurer la qualité des données, développer la documentation.

Des outils peuvent être utilisés pour une meilleure documentation de la valeur (ex: Business Model Canvas) et des risques (ex: Factsheet IBM).

ACTEURS CONCERNÉS

Cette phase nécessite une équipe **pluridisciplinaire** dotée de l'éventail de compétences fonctionnelles, avec des experts du domaine pour influer sur les choix de conception pertinents.

Les responsables s 'engagent et mobilisent leurs équipes : par exemple le sponsor et les acteurs métiers, ainsi que les équipes Data/IA. Dès le lancement il pourrait y avoir un choix éthique sur le lancement où non du projet d'IA et des orientations particulières.

Les partenaires **externes** s'il y a lieu doivent être intégrés à la décision d'engagement des ressources.

La responsabilité des systèmes repose notamment sur la **responsabilité individuelle** des membres de l'équipe.

POINTS DE VIGILANCE

La phase peut correspondre à plusieurs cycles d'idéation avant lancement : une difficulté est de savoir quitter le mode idéation pour commencer à proprement dit la qualification. Elle doit permettre d'alimenter une **cartographie** globale des cas d'usages d'IA lancés qui n'existent pas toujours.

Il est difficile de qualifier (financièrement) la valeur de la dimension responsable. L'échange entre acteurs techniques et marketing peut être complexifié par le manque de **vocabulaire** commun, d'objectif commun et de métriques communes.

L'évaluation éthique s'ajoute à des activités existantes d'évaluation des risques sur les données personnelles ou sur la sécurité et doit trouver sa place dans l'organisation, sous la forme de règles imposées, sous réserves que ces règles aient du sens dans tous les contextes, ou sous la forme de recommandations, appliquées de manière volontaire dans les cas où elles s'avèrent pertinentes.

Il est difficile de trouver la réponse **proportionnelle** au risque identifié. La question de l'usage de composants réutilisables, développés en interne ou open-source, versus le développement d'une solution ad-hoc, doit être évaluée.

CONCEPTS CLÉS

Analyse de risques et d'impacts | Contraintes opérationnelles | Organisation de la Gouvernance

LES BONNES PRATIQUES IDENTIFIÉES

En s'inspirant ***des bonnes pratiques extraites de la certification IA proposée par le LNE :***

Les éléments **d'entrée et sortie** à documenter pour la phase de qualification sont à minima :

Pour les éléments d'entrée :

- Les exigences, attentes, bénéfices attendus des personnes cibles pour le système à base d'IA,
- Les exigences réglementaires applicables dans les zones concernées,
- Les exigences issues d'activités et/ou usages similaires (normes, règles de l'art, retours d'expérience).

  Le client peut être spécifique dans le cas où la fonctionnalité d'IA est développée pour un acteur en particulier ou bien générique dans le cas où il est prévu que la fonctionnalité d'IA soit distribuée largement.

Pour les éléments de sortie :

- Les spécifications de la fonctionnalité d'IA,
- Les perspectives en matière d'opérationnalisation, de passage à l'échelle et de réutilisation éventuelle,
- Les exigences concernant la documentation associée,

  Par exemple, doivent être définis les besoins en documentation utilisateur ou concernant la fiche produit. Peuvent également être définies que les plans de test permettant la vérification des exigences relatives à la fonctionnalité d'IA doivent être associés à la documentation utilisateur.
- Les besoins de communication avec le client et les utilisateurs,
- Une analyse de risques avec la précision des éléments suivants:
  - Les modes de défaillance prévisibles de la fonctionnalité d'IA avec les scénarios qui pourraient conduire à une défaillance et à un préjudice
  - Les implications sociétales et environnementales d'une défaillance prévisible du produit, d'une mauvaise utilisation ou d'une attaque malveillante
  - Les utilisations potentielles non planifiées de la fonctionnalité d'IA et du produit

"Il est indispensable d'ajouter la partie accès aux données dans la qualification des projets d'IA en plus des critères traditionnels de qualification projet : ROI, alignement stratégique, ressources et pour intégrer la dimension responsable il est important de mettre en avant les bénéfices attendus et le niveau de risque pour les personnes."

Émilie Sirvent-Hien - Responsable AI Program Manager - Orange

PHASE 3

IMPACT AI

DÉVELOPPEMENT D'UN SYSTÈME À BASE D'IA

OBJECTIF

Pour développer un système à base d'IA de manière responsable il s'agit notamment :

- De concevoir la fonctionnalité d'IA pour atténuer les impacts négatifs potentiels identifiés lors de la phase de qualification (sur les personnes, l'environnement, la société, etc.),
- D'**évaluer les données** et les résultats de la fonctionnalité d'IA pour minimiser le risque de biais et d'atteinte à l'équité, et d'établir pour cela les métriques clé nécessaires,
- D'incorporer des fonctions permettant autant que de besoin, le contrôle de la fonctionnalité par l'être humain,
- De prendre des mesures pour protéger les données et la fonctionnalité d'IA,
- De développer la documentation tout au long du cycle de développement pour assurer la transparence.

PROCESSUS - OUTILS - PILOTAGE

Les standards de développement à établir, les pratiques à mettre en œuvre dans ce cadre, et l'outillage à définir doivent :

- Assurer (au mieux) le suivi des sources de données, leur validation, et leurs utilisations (autorisées),
- Permettre la mesure de la cohérence des résultats de la fonctionnalité d'IA avec l'**objectif** retenu (et d'éviter/d'identifier toute dérive du modèle, tout baisse de la qualité des données), la mesure et l'analyse des biais et l'application de techniques d'atténuation, les tests d'expédition (release) et critères associés ainsi que le respect continu des mesures, tests et critères d'équité pour la production et les usages,
- Garantir l'intégration de **choix de conception** pour atténuer ou contrôler les impacts négatifs identifiés (ex. à la suite d'une défaillance, d'une utilisation non planifiée, d'une attaque ou simplement d'un effet secondaire d'une utilisation normale de la fonctionnalité),
- Garantir l'intégration de choix de conception essentiels pour une utilisation appropriée de la fonctionnalité d'IA, une collecte de données légitime et transparente, ainsi que :
  - La sécurité (ex. identification et traitement des vulnérabilités de sécurité et des vecteurs d'attaques possibles (empoisonnement des données ou autres modèles d'adversaires valides), chiffrement des données, protection de la propriété intellectuelle de la fonctionnalité d'IA, etc.),
  - La confidentialité des données sensibles et ou le respect de la vie privée (ex. anonymisation/pseudonymisation, non-divulgation par inadvertance d'informations sensibles ou privées pendant l'utilisation de la fonctionnalité),
  - Et la conformité avec la législation en vigueur (ex. pour les usages dit sensibles).
- Préciser quelles caractéristiques de la fonctionnalité d'IA et du produit associée garantiront des expériences inclusives,
- Prendre en compte les décisions ou les fonctions qui nécessitent une supervision humaine en tant que composante essentielle de la fonctionnalité d'IA et définir pour cela les mécanismes (ex., l'interprétabilité) pour favoriser la compréhension de ladite fonctionnalité d'IA par l'utilisateur final afin de permettre une vérification, une surveillance et une intervention humaine en continu,
- Définir les canaux utilisés pour recueillir des commentaires en direct,
- Documenter de manière cohérente:
  - Les choix de conception et de développement précédents, les raisonnements et les hypothèses afférentes,
  - Les types de modèles, d'outils ou de techniques pour documenter le comportement de la fonctionnalité d'IA.

CONCEPTS CLÉS

Data scientist | Choix et mise en oeuvre des Métriques | Compromis

ACTEURS CONCERNÉS

Les acteurs de cette phase sont multiples, pilotés par le coordinateur de projet :

- Acteurs métier concernés
- Data Scientists en charge de l'entraînement du/des modèle(s)
- Développeurs (Software Engineering) (pour le front end / back end)
- UX Designer (pour l'expérience utilisateur)
- Future équipe de mise en production et de maintenance (dont les ML Engineers).

POINTS DE VIGILANCE

- Manque d'**implication** et de responsabilisation des demandeurs/futurs **utilisateurs** (ex: importance de la labellisation des jeux de données par des sachants du métier)
- Manque d'**expertise métier** lors du développement (ex: le modèle émule des règles métiers et doit pouvoir être jugé sur cette base, au delà des métriques statistiques) et risque de perte de sens (ex: paradoxe de simpson qui va corréler des phénomènes distincts)
- Solution qui ne sera pas correctement utilisée car l'expérience utilisateur n'a pas été suffisamment travaillée: l'adhésion des utilisateurs au mode d'utilisation du système est indispensable (ex: une boîte noire, aussi performante soit-elle, ne conviendra pas à certains usages, où la confiance dans le résultat passe par son explication possible)
- Manque de **diversité** dans l'équipe de conception (ex: risque de biais culturels / 'angles morts' lors des tests).

LES BONNES PRATIQUES IDENTIFIÉES

- Définition claire de l'**objectif** du modèle (incluant un score de référence, à améliorer)
- Fiche d'identité du système d'IA, permettant de le détailler depuis sa conception jusqu'à son utilisation dans un document unique pour rassembler les informations techniques concernant le modèle (renouvelé ou complété à chaque itération), pour consultation par tous et qui suit la vie du modèle (ex: ***Rapport Éthique Zelros***, Fiche produit Maif, Dataiku feature generation documentation)
- Implication de l'ensemble des acteurs lors du développement: L'utilisateur final - qui fera partie des responsables du modèle une fois en production - doit être représenté lors du développement
- Partir d'une solution sans apprentissage supervisé (ML), pour disposer d'un état de référence, que l'utilisation du ML doit permettre d'améliorer
- Acculturation des acteurs aux forces et faiblesses des algorithmes d'IA
- Développer les outils de **communication** entre l'ensemble des acteurs (prescripteurs métiers, IT, Data Science, utilisateurs métiers, auditeurs...).

"La phase de développement est une phase délicate du cycle: il faut trouver le bon compromis entre attente utilisateur et performance raisonnable, ménageant la transparence et l'explicabilité.

L'IA responsable nécessite pour cela de garder l'humain au cœur du développement : communication entre futurs utilisateurs et concepteurs, réflexion sur les biais possibles, humilité quant à la performance imaginée et anticipation des impacts liés à l'usage.

C'est une phase passionnante !"

Antoine de Langlois, Responsible AI Lead @ Zelros

PHASE 4

IMPACT AI

L'IA RESPONSABLE EN PHASE DE PRODUCTION

OBJECTIF

La mise en production est un processus qui doit être anticipé dès la phase de conception. Il s'agit principalement ici d'assurer la **continuité** en ce qui concerne les principes éthiques et valeurs de l'organisation et la responsabilité. L'idée est d'établir dans l'organisation une politique standardisée de gestion des systèmes d'IA à l'image des politiques de sécurité et de protection des données personnelles existantes. Ceci rentre dans le périmètre de la gouvernance de tels systèmes.

PROCESSUS - OUTILS PILOTAGE

Les outils habituels de la mise en production et du MLOps peuvent ici être utiles :

- Monitoring de la performance
- Plan de continuité et de reprise de projet

Pour garantir une mise en production responsable, le suivi des **métriques** établies, du caractère toujours éthique des données s'appuiera également sur des questionnaires auprès des équipes conceptrices. Les **tableaux de bord** résultants et ces questionnaires pilotent l'IA vers sa mise en production.
Ils permettent de suivre les risques et valeurs pris en compte lors de la conception et de contrôler l'efficacité des techniques mises en œuvre.

ACTEURS CONCERNÉS

- Machine Learning Engineer (chargé de la mise en production en partant du modèle entraîné)
- Chefs de projets, équipes dirigeantes
- Tout acteur chargé d'assurer les liens entre équipes techniques et utilisation commerciale
- Équipes Devops et de maintenance

POINTS DE VIGILANCE

Nous identifions 5 **sources majeures de risque** dans la phase de production :

- Le risque d'une dérive ou d'une baisse de qualité des données, ou d'une évolution du comportement des consommateurs (on peut penser ici au risque sur un modèle de détection de fraude, si le comportement des fraudeurs s'ajuste),
- Un risque technique, avec la possibilité d'un arrêt dans la chaîne de transmission des données, qui peut amener à un arrêt du service ou une baisse de performance,
- Un risque de sécurité, impliquant par exemple une fuite de données sensibles ou personnelles ou une attaque du modèle,
- Un risque réglementaire : une fois le projet mis en production, l'importance de se conformer à la réglementation en vigueur, qui peut évoluer, est décuplée,
- Enfin, le risque sans doute le plus important est un risque d'image publique, par exemple si des biais algorithmiques sont découverts par le public avant d'avoir pu être identifiés par l'entreprise.

Outre les risques, un enjeu de la mise en production est de faire passer les critères éthiques du projet à l'application de **grande échelle**. Quelques exemples de dilemmes :

- *Responsabilité* : comment anticiper toutes les conséquences indésirables du déploiement de l'IA, alors que le modèle n'a pas été entraîné sur tous les scenarii possibles ?
- *Transparence et équité* : compréhension des outils data-science par le métier (tous ne sont pas présents aux sessions de formation ; entre vulgarisation et perte de l'information sur l'IA)
- *Critères éthiques plus impalpables*, comme le « bien-être » : comment les cadrer et garantir leurs applications par des outils ?

CONCEPTS CLÉS

ML Engineer | Mise en conditions opérationnelles | Passage à l'échelle | Contrôles de sécurité

LES BONNES PRATIQUES IDENTIFIÉES

Pour détecter en amont et pouvoir remédier à une baisse de qualité du modèle (dû à une dérive des données ou un problème technique dans la chaîne de transmission), il est possible de mettre en place des dashboards de suivi, permettant un monitoring du modèle.

Ce monitoring sera plus efficace s'il concerne plusieurs indicateurs, qui doivent être bien choisis.

Ces indicateurs peuvent également être multipliés sur différentes populations, particulièrement si certaines populations sont identifiées comme potentiellement à risque.

On peut ainsi suivre un indicateur sur le groupe des femmes de plus de 60 ans, ou celui des habitants d'un département, etc.

Afin de pallier les risques techniques, nous conseillons d'avoir des plans de continuité et de récupération des données ou du modèle, ainsi que des sauvegardes régulières.

Pour faire face au risque de vol de données, mettre en place des mesures de sécurité à la mesure du risque, tel que limiter l'accès aux données aux personnes essentielles et habilitées.

Quant au risque sur l'image publique, il est plus facile de le prévenir que de le guérir.

Pour cela, on peut par exemple adopter un code d'éthique en interne, qui permettra d'une part de diminuer les risques et d'autre part de montrer la bonne volonté de l'entreprise sur ces sujets.

Il est aussi possible d'organiser des « algorithmic bias bounty challenges », consistant à faire appel à la communauté pour rechercher et identifier des biais possibles dans les algorithmes, pour les corriger avant la mise en production.

Outre l'évitement des risques, l'utilisateur peut garantir l'application de ses critères éthiques face à certains **dilemmes** :

- **Transparence et équité**

Préparant la compréhension de l'IA par un profil moins technique, la solution peut être auditée par un utilisateur-métier. Les data-scientists peuvent partir de ces audits internes pour organiser des formations plus accessibles (comment utiliser les paramètres, interpréter l'explicabilité ou telle métrique d'équité…).

- **Critères éthiques plus impalpables**

Un premier enjeu est de faire le pont entre une perception vague, du « bien-être » par exemple, et des directions plus précises. Trouver des cadres théoriques en lien avec les politiques publiques (ONU, UE, nationale) peut être une voie. Par exemple, le principe de « bien-être » peut être défini à travers les 17 Objectifs pour le Développement Durable de l'ONU pour 2030 (Cf. livre blanc de Nicolas Meric et Thomas Souverain « State of Ethical AI in 2021 ») : accès à une éducation de qualité, consommation énergétique limitée…

Puis il s'agit de les traduire en applications concrètes, en s'assurant que ces directions sont respectées dans l'IA mise en production (mesures de l'entreprise, librairies de code, applications IA orientant les clients vers des solutions écologiques…).

Organiser une boucle de feedback de validation itérative. PDCA (Plan, Do, Check, Adjust)

Organiser le monitoring et loguer les alertes de défaut (exemple dérive des données…).

Organiser une boucle de feedback par les utilisateurs (avec remontée automatique) pour amélioration au fur et à mesure.

"Si la data responsable se conçoit dès l'origine du projet, avec le choix du cas d'usage, des données et de la cible, la phase de production n'est cependant pas à négliger car c'est lors de cette phase que le projet rencontre réellement ses utilisateurs et peut se mettre à avoir un impact sur eux et sur son environnement."

Irene Balmes, Manager associée - Lead Data Scientist Kynapse

PHASE 5

IMPACT AI

PILOTAGE DES USAGES RESPONSABLES D'UN SYSTÈME À BASE D'IA

OBJECTIF

Le pilotage des usages responsables d'un système à base d'IA est l'une des phases critiques du cycle de vie d'un système à base d'IA. C'est lors de ses usages qu'un système peut produire des **externalités négatives** (e.g. résultats opaques et/ou discriminatoires) avec des impacts financiers et réglementaires considérables pour les organisations (e.g. non-conformités avec les cadres réglementaires actuels [RGDP] et à venir [AI Act]). Cette fiche pratique recense des processus et outils utilisés au sein des organisations pour piloter les usages des systèmes à base d'IA après déploiement et assurer un usage responsable.

PROCESSUS - OUTILS - PILOTAGE

- Outils et processus de gestion des opérations ML (MLOps) : Une fois le modèle entraîné et confirmé conforme du point de vue de la performance (sur un jeu de test différent de celui de l'entraînement) et de nos critères d'équité (IA responsable), il peut être mis en production. Une fois en production, il convient de s'assurer qu'il continue de se comporter conformément aux attentes grâce à des métriques clés comme le ***drift***. Comme un rapport généré à l'issue de l'entraînement du modèle (ex: ***Rapport Éthique Zelros*** ou **Model Document Generator de Dataiku**) peut être intéressant à conserver comme une sorte de **fiche d'identité du modèle**, un rapport de monitoring sur une période donnée peut être un support utile pour faire ce contrôle. Ce rapport devra permettre d'apprécier comment s'est comporté le modèle sur la période, par exemple en observant:
  - Est-ce que les valeurs utilisées à la prédiction pour les variables du modèle correspondent-elles à celles observées lors de l'entraînement?
  - Peut-on qualifier un "Data Drift" (changement dans la distribution des données entre l'entraînement et la production) ou un "Concept Drift" (les règles métiers déduites par le modèle lors de l'entraînement ne sont plus d'actualité)?
  - Est-ce que le modèle est toujours autant utilisé **dans le temps**?
  - Est-ce que les prédictions prennent de plus en plus de temps?
  - Quelle est la consommation carbone associée à l'ensemble des prédictions sur la période ?

En fonction des résultats, autrement dit l'identification d'un drift, il sera important de ***ré-entraîner le modèle*** afin de les traiter dans la durée et préserver la qualité des prédictions du modèle.

- Outils et processus de Gouvernance : la gouvernance de projets doit elle aussi se mettre à jour à l'heure de l'IA. Alors que de nombreux concepts et cadres émergent au sein des organisations (DevOps, MLOps, IA Éthique et Responsable) pour piloter les projets d'Analytique et d'IA, la gouvernance IA joue un rôle clé afin de les standardiser et faciliter le passage à l'échelle. Sans gouvernance, les risques liés à l'IA sont traités de manière hétérogène et créent des risques opérationnels forts (Par exemple, la plateforme de data science Dataiku propose un nouvel espace centralisé, appelé "Govern" (accéder à une vidéo de présentation en Anglais ***ici***), afin de remplir ces objectifs. Les difficultés rencontrées ?
- De manière générale, les méthodes et outils de pilotages de projets d'IA commencent à être identifiés par les organisations mais sont encore trop peu utilisés
- Manque d'adhésion des utilisateurs: usage inadapté des résultats du modèle/système d'IA, défaut de confiance dans les résultats...
- Les talents ne sont pas toujours bien formés lorsqu'il s'agit:
  - d'analyser la performance du modèle selon des objectifs d'équité
  - d'identifier les drifts (data/modèles) nécessaires et suffisants pour une application donnée
  - de réagir aux drifts (fréquence à définir - hebdomadaire, mensuel,...)
- Les enjeux de cybersécurité spécifiques à l'IA (par exemple l'attaque des modèles) sont encore peu abordés au sein des entreprises.

CONCEPTS CLÉS

Utilisateurs | Communication | Détection et correction des dérives

POINTS DE VIGILANCE

- De manière générale, les méthodes et outils de pilotages de projets d'IA commencent à être identifiés par les organisations mais sont encore trop peu utilisés
- Manque d'adhésion des utilisateurs: usage inadapté des résultats du modèle/système d'IA, défaut de confiance dans les résultats...
- Les talents ne sont pas toujours bien formés lorsqu'il s'agit:
  - d'analyser la performance du modèle selon des objectifs d'équité
  - d'identifier les drifts (data/modèles) nécessaires et suffisants pour une application donnée
  - de réagir aux drifts (fréquence à définir - hebdomadaire, mensuel,...)
- Les enjeux de cybersécurité spécifiques à l'IA (par exemple l'attaque des modèles) sont encore peu abordés au sein des entreprises.

ACTEURS CONCERNÉS

- Utilisateurs
- Équipe opérationnalisation et maintenance
- Équipes de risques, compliance et conformité
- Chefs de projets, équipes dirigeantes.

LES BONNES PRATIQUES IDENTIFIÉES

- Importance d'**anticiper** en amont les risques en amont et de construire des documents pertinents pour chaque étape du cycle de vie d'un projet d'IA. Concernant l'utilisation d'un projet d'IA, il est important d'utiliser
  - un document référence guidant les équipes dans le pilotage de projets d'IA
  - des processus standardisés afin de faciliter l'identification et le traitement des risques
  - des outils de gouvernance permettant de faciliter la documentation et le reporting d'un projet d'IA.
- Cas par cas : il est important de considérer le pilotage d'un projet d'IA en fonction de son cas d'usage. En fonction des données/techniques utilisées ou tout simplement des utilisateurs ou de l'industrie concernée, le **pilotage** devrait être renforcé.
- Suivi de l'expérience utilisateur, via des ateliers par exemple: un produit, aussi performant soit-il, ne sera utilisé que s'il est adopté par les utilisateurs (utilisation simple, intégrée dans l'environnement habituel de l'usager). Adaptation en conséquence de l'UX / ergonomie.
- Possibilité de prise en compte des **retours utilisateurs**, soit pour amélioration en continu (reinforcement learning) ou simplement pour une revue régulière de la solution par des humains.

"Alors qu'un cadre réglementaire pour l'IA voit le jour au sein de l'Union Européenne, il est fondamental pour les organisations de pouvoir communiquer avec leurs parties prenantes sur le déploiement et l'adoption d'un projet d'IA. La standardisation et la documentation de la fin du cycle de vie d'un projet IA permettent de répondre facilement à ces enjeux tout en testant sa capacité de mise en conformité avec le futur cadre réglementaire»

Paul-Marie Carfantan, AI Governance Solution Manager - Dataiku

LEXIQUE

LEXIQUE PROPOSÉ POUR APPUYER NOTRE APPROCHE DE DÉVELOPPEMENT RESPONSABLE DES SYSTÈMES D'IA.

ANALYSE DE RISQUES ET D'IMPACTS

Une analyse de risque a pour but l'identification de différents types de risques associés à différents scénarios de défaillance potentiels. Le risque est la combinaison de la probabilité d'apparition et de la gravité. L'objectif principal est de fournir la meilleure sélection de moyens permettant de contrôler ou d'éliminer les risques identifiés.

Pour les systèmes d'IA des nouveaux types de risques sont à prendre en compte (exemple: perte de contrôle humain, manque de fiabilité, discrimination...). Une analyse d'impact relative à la protection des données est une procédure de contrôle parfois obligatoire pour déterminer si un traitement de données personnelles est susceptible de présenter un risque important pour les droits et libertés des individus concernés par ce même traitement. Par extension, une analyse d'impact éthique pour un système d'IA évalue l'impact pour les personnes et les collectifs concernés par le système directement ou indirectement.

BESOINS MÉTIER

Les besoins métiers correspondent aux besoins des opérationnels qui utilisent ou exploitent les systèmes existants (avec ou sans IA), qui peuvent rencontrer des difficultés (de coût, performance, manque d'utilisabilité...) et donc avoir des besoins d'innovation. Recueillir le besoin signifie amener le donneur d'ordre métier à formuler une expression claire de sa demande.

C'est un travail qui relève autant du design stratégique que de l'ergonomie. Et par extension, on est également dans les sphères du design d'expérience utilisateur. Plusieurs outils et méthodes peuvent être mises en œuvre, des plus formelles aux plus informelles : l'enquête, l'analyse quantitative et qualitative, le retour d'expérience...

CHOIX ET MISE EN ŒUVRE DES MÉTRIQUES

Une métrique d'évaluation quantifie la performance d'un modèle prédictif. Le choix de la bonne métrique est donc crucial lors de l'évaluation des modèles de Machine Learning, et la qualité d'un modèle de classification dépend directement de la métrique utilisée pour l'évaluer. A noter qu'il existe plusieurs types de métrique:

- de robustesse ou performance par exemple courbe ROC, taux de faux positifs, négatifs...
- d'équité: métrique de parité statistiques, asymétrie du taux d'erreur, calibration...

COMMUNICATION

La communication sur les systèmes d'intelligence artificielle fait partie du thème de la transparence. Il s'agit de communiquer aux différents types d'utilisateur la finalité du système, ses limites, comment l'utiliser, pourquoi de tels résultats et également s'assurer que les différents utilisateurs ont compris les éléments fournis.

COMPROMIS

En matière d'intelligence artificielle, il y a souvent des choix à faire (par exemple de métrique d'évaluation à utiliser, de seuils de décision, données à utiliser ou non...). Il y a souvent des objectifs contradictoires et les choix éthiques réalisés sont donc le résultat d'un compromis réalisé par un décideur ou un collectif.

CONTRAINTES OPÉRATIONNELLES

Les contraintes opérationnelles sont importantes à prendre en compte lors de la conception d'un système d'intelligence artificielle pour s'assurer du succès lors du déploiement. Les contraintes peuvent être de natures très variées (économiques,d'utilisabilité, temporelles, de sécurité...)

CONTRÔLES DE SÉCURITÉ

Lors du déploiement des systèmes d'IA il est important de réaliser des contrôles de sécurité pour évaluer les risques et les corriger. Il peut y avoir des risques sur l'accès aux données et également des nouveaux risques liés à de nouveaux types d'attaques sur les modèles d'IA.

DATA SCIENTIST

Spécialiste des statistiques, de l'informatique et du marketing, le Data Scientist recueille, traite, analyse et fait parler les données qui peuvent être massives dans le but de réaliser un système qui répond à un besoin métier (en général pour améliorer les performances d'une entreprise.)

DÉTECTION ET CORRECTION DES DÉRIVES

La dérive du modèle (également connue sous le nom de dégradation du modèle) fait référence à la dégradation du pouvoir de prédiction d'un modèle en raison de changements dans l'environnement, et donc dans les relations entre les variables.

Par exemple, des changements dans la présentation

LEXIQUE

LEXIQUE PROPOSÉ POUR APPUYER NOTRE APPROCHE DE DÉVELOPPEMENT RESPONSABLE DES SYSTÈMES D'IA.

des e-mails de spam entraîneraient la dégradation des modèles de détection des fraudes créés il y a plusieurs années.

Il existe trois principaux types de dérive de modèle :

- La dérive des concepts
- La dérive des données
- Les changements de données en amont

La dérive du concept est un type de dérive du modèle où les propriétés de la variable dépendante changent. La dégradation de modèles de détection de fraude évoquée ci-dessus est un exemple de dérive conceptuelle, où la classification de ce qui est «frauduleux» change.

La dérive des données est un type de dérive du modèle où les propriétés de la ou des variables indépendantes changent. Les exemples de dérive des données comprennent les changements dans les données dus à la saisonnalité, les changements dans les préférences des consommateurs, l'ajout de nouveaux produits, etc...

Les changements de données en amont font référence aux changements de données opérationnelles dans le pipeline de données. Par exemple, lorsqu'une fonctionnalité n'est plus générée, ce qui entraîne des valeurs manquantes. Un autre exemple est un changement de mesure (par exemple, des miles aux kilomètres).

MISE EN CONDITIONS OPÉRATIONNELLES

Une fois le système d'IA validé au niveau de ses performances, le déploiement correspondant à l'étape de mise en production en conditions opérationnelles, c'est-à dire dans les mains des utilisateurs cibles, avec des données et un environnement de production. La validation de cette étape est très importante car il y a souvent un décalage entre l'environnement et les données de test et la production. A noter qu'après le premier déploiement il est important d'anticiper le système dans la durée et notamment le maintien en conditions opérationnels avec un niveau de performance stable.

ML ENGINEER

En plus de développer des algorithmes, le Machine Learning Engineer est aussi responsable de la mise en production (industrialisation) des modèles d'intelligence artificielle. Cet expert a donc la double casquette de scientifique des données et de développeur logiciel.

PARTIES PRENANTES DIVERSES

Les parties prenantes correspondent à l'ensemble des personnes physiques et morales qui sont concernées par un système et qui influencent les décisions prises lors de la conception ou exploitation du système.

Elles se distinguent notamment par leur pouvoir basé sur ce qu'elles apportent, ainsi que sur ce qu'elles en attendent. En matière d'éthique, il est important de considérer la diversité des parties prenantes pour s'assurer d'une représentativité des utilisateurs du système après déploiement et d'éviter des biais cognitifs dans la conception.

ORGANISATION DE LA GOUVERNANCE

La gouvernance d'une entreprise fournit un cadre pour surveiller les actions stratégiques menées, la définition et la bonne utilisation des moyens pour un niveau de performance attendu. Elle définit qui contrôle quoi et comment ce contrôle s'exerce.

Pour cela, elle fixe des règles, des pratiques à suivre et des codes de conduite selon les situations rencontrées. En matière d'intelligence artificielle une gouvernance de l'IA avec des nouveaux rôles est à créer par les entreprises (par exemple pour gérer les données, les modèles d'IA et leur mise en production, les comités éthiques...)

UTILISATEURS

Le terme utilisateur est employé pour désigner une personne qui utilise un système informatisé (ordinateur, robot par exemple) mais qui n'est pas nécessairement informaticien (par opposition au développeur ou data scientist par exemple). Les systèmes d'intelligence artificielle peuvent être destinés à des profils d'utilisateurs très variés, d'où le besoin d'inclure dans la conception les utilisateurs y compris pour la validation des systèmes.

VALEURS ÉTHIQUES

Dans une perspective éthique, une valeur représente ce qui inspire, motive et guide nos décisions et nos actions dans nos rapports avec autrui. Elle constitue la fin visée par la décision ou l'action envisagée et se traduit verbalement comme raison d'agir et comme sens de l'action. En matière d'intelligence artificielle on peut citer des exemples de valeurs éthiques comme la transparence, l'équité par exemple.

PASSAGE À L'ÉCHELLE

Le passage à l'échelle, anglicisme pour le redimensionnement et la mise à l'échelle, est la faculté qu'a un système à pouvoir changer de taille ou de volume selon les besoins des utilisateurs. La disponibilité et la latence font partie des critères de qualité d'un système.

CONCLUSION
ET PROCHAINES ÉTAPES

Une démarche responsable est un processus en amélioration continue qui sait rester modeste et cherche à s'améliorer en permanence.

Plus précisément, une telle démarche encourage la standardisation de cadres opérationnels et de valeurs par les chefs de projets data afin de systématiser la sélection, développement et déploiement de projets d'Analytique et d'IA. Les capacités et les savoir-faire développés facilitent la compréhension de l'ensemble des projets au sein d'une organisation.

L'utilisation d'outils adaptés permet à l'ensemble des acteurs d'un projet d'IA, qu'ils soient techniques ou non, d' accéder aux informations nécessaires et de les utiliser selon leurs objectifs et besoins.

Le pilotage sur l'ensemble du cycle de vie d'un projet d'IA se retrouve facilité car l'ensemble des projets sont qualifiés, documentés et prêts a être audités par les régulateurs ou par un organisme certificateur pour faire reconnaître les efforts entrepris.

Cette régulation sera le thème du prochain cycle de travail du collectif Impact AI, IA Responsable en 2022.

À bientôt !

INTERVENANTS

Coordination du groupe de travail sur les fiches pratiques:

EMILIE SIRVENT HIEN - **ORANGE** | NAIRA HAMBARDZUMYAN - **DELOITTE** | ROXANA RUGINA - **IMPACT AI**

Membres du groupe de travail IA Responsable ayant participé aux différents ateliers de partage :

MARCIN DETYNIECKY	**AXA**
CHAOUKI BOUTHAROUITE	**AXA**
GWENDAL BIHAN	**AXIONABLE**
NICOLAS BLANC	**CFE CGC**
ALDRICK ZAPPELLINI	**CREDIT AGRICOLE**
PAUL-MARIE CARFANTAN	**DATAIKU**
GREGORY ABISROR	**DELOITTE**
YVES YOTA TCHOFFO	**DELOITTE**
NAIRA HAMBARDZUMYAN	**DELOITTE**
AXEL CYPEL	**RESPONSABLE PROJETS IA**
ISMAEL AL-AMOUDI	**GRENOBLE MANAGEMENT SCHOOL**
FLORENCE PEUTIN	**INTEL**
FLORENCE PICARD	**INSTITUT DES ACTUAIRES**
IRENE BALMES	**KYNAPSE**
ANNE-MARIE JONQUIERE	**LE CERCLE INTERELLES**
SANDRINE BORTON	**LE CERCLE INTERELLES**
SWEN RIBEIRO	**LNE**
BARBOSA VIRGINIE	**LNE**
AVRIN GUILLAUME	**LNE**
NICOLAS MARESCAUX	**MACIF**
OLIVIER BAES	**MAIF**
DANIEL BARTOLO	**MAIF**
BERNARD OURGHANLIAN	**MICROSOFT**
PHILIPPE BERAUD	**MICROSOFT**
ENERIC LOPEZ	**MICROSOFT**
GUILLAUME DELANDTSHEER	**NETAPP**
SIMON DECARPENTRIES	**NETAPP**
GUY MAMOU-MANI	**OPEN**
EMILIE SIRVENT HIEN	**ORANGE**
JEAN-DAVID BENASSOULI	**PWC**
CYRIL JACQUET	**PWC**
MARC DAMEZ FONTAINE	**PWC**
CLAUDE LE PAPE GARDEUX	**SCHNEIDER ELECTRIC**
VALERIE NOELLE KODJO DIOP	**SOCIÉTÉ GÉNÉRALE**
ERIC BONIFACE	**SUBSTRA FOUNDATION**
BLOCH EMMANUEL	**THALES**
ANTOINE DE LANGLOIS	**ZELROS**

POUR PLUS D'INFORMATIONS *WWW.IMPACT-AI.FR*