numeum

BEHIND THE CODES AND THE DATA

GUIDE PRATIQUE POUR DES IA ÉTHIQUES

[Version 1 — Septembre 2021]

SOMMAIRE

LA DÉMARCHE

- L'ambition de Numeum et de ses partenaires a été de traduire des **principes éthiques** généraux, issus de travaux existants, en **méthodes pratiques.**
- **7 ateliers** d'intelligence collective, rassemblant plus de **350 participants**, ont été conduits entre le 10 novembre et le 22 décembre 2020. Ils ont permis de rassembler les **propositions** et les **recommandations** des contributeurs pour chaque thème identifié.
- La consolidation et la formalisation de ces travaux se sont poursuivies de janvier à avril 2021, encadrées par un **Comité de relecture** composé de **partenaires académiques** et de **professionnels**.
- **Avertissement :** la réflexion s'est volontairement concentrée sur des systèmes d'IA simples, permettant de facilement se projeter sur des cas d'application concrets. Les systèmes d'IA complexes, totalement autonomes dans la prise de décision - tels que les véhicules et autres robots sans interventions humaines - ont donc été exclus des travaux.
- Ce **Guide pratique** a été officiellement présenté aux pouvoirs publics ; il constitue, par ailleurs, le socle du **Manifeste pour des IA éthiques.** Retrouvez tous ces éléments sur le site Ethical AI, dédié à la démarche : ai-ethical.com.
- Des **mises à jour régulières** permettront de compléter et d'actualiser ce document au gré des **évolutions technologiques** ou des **nouvelles exigences** (réglementaires, sociétales, économiques...).

Renaud VEDEL, *Coordonnateur de la Stratégie nationale pour l'intelligence artificielle (CSN-IA)*

Katya LAINÉ, *Administratrice et Présidente du Comité Innovation & Technologies de Numeum*

PAS DE NUMÉRIQUE RESPONSABLE SANS IA ÉTHIQUES

Porteuse de progrès inédits, l'intelligence artificielle (IA) se diffuse dans toutes les sphères de notre vie quotidienne. Ce formidable essor pose notamment **la question légitime de la confiance** que l'être humain peut avoir dans ces systèmes.
Pour y répondre, il est nécessaire que les solutions développées respectent les droits fondamentaux défendus, en particulier, par la France et l'Union européenne.
Pour s'épanouir pleinement, l'IA doit donc être éthique : c'est une des conditions essentielles qui lui permettront de tenir toutes ses promesses **au bénéfice du plus grand nombre**. Mais la tâche est ardue quand il s'agit de passer de la théorie à la pratique. C'est donc avec la même détermination que celle qui les motive à faire progresser la technologie que des spécialistes de l'IA se sont mobilisés pour définir un cadre opérationnel permettant de créer et de diffuser des IA éthiques.

Numeum a fédéré un réseau de partenaires issus des différents univers de l'IA – monde académique, pouvoirs publics, entreprises, milieu associatif et société civile. Les très nombreux échanges et partages d'expérience ont permis d'aboutir à ce guide pratique qui propose **une méthode utile pour opérationnaliser les principes éthiques majeurs**, lors de la conception, le développement et le déploiement de solutions d'IA.
Cet outil constitue ainsi un **code de conduite volontaire** pour développer des IA de confiance comme l'encourage la Commission européenne[1]. Fort du crédit apporté par les partenaires engagés dans la démarche, il a vocation à devenir un outil de référence en France et au-delà.

Cette initiative pour doter les IA de qualités éthiques tangibles enrichit notre démarche de développement d'un numérique plus responsable. Elle constitue une étape clé dans l'anticipation et la préparation au respect des principes de la future réglementation générale pour l'IA récemment proposée par la Commission européenne.

1. Voir https://ec.europa.eu/commission/presscorner/detail/fr/QANDA_21_1683

LES PARTENAIRES

Cliquer sur une photo pour lancer la vidéo associée

Laurence DEVILLERS, *Professeure en IA à l'Institut Data IA*

Charles BOUVEYRON, *Directeur de l'Institut 3IA Côte d'Azur*

Roxana RUGINA, *Secrétaire Générale d'Impact AI*

Sophie VIGER, *Directrice Générale de l'école 42*

Tawhid CHTIOUI, *Président-fondateur & Dean d'aivancity*

Magali BARNOIN, *Animatrice du numérique, Data & IA de Telecom Valley*

Antoine TROTET, *Chef du service Révolution Numérique de la région Grand Est*

Gaëlle PINSON, *Directrice Générale du Hub France IA*

Nicolas VIALLET, *Directeur Opérationnel ANITI, Université de Toulouse*

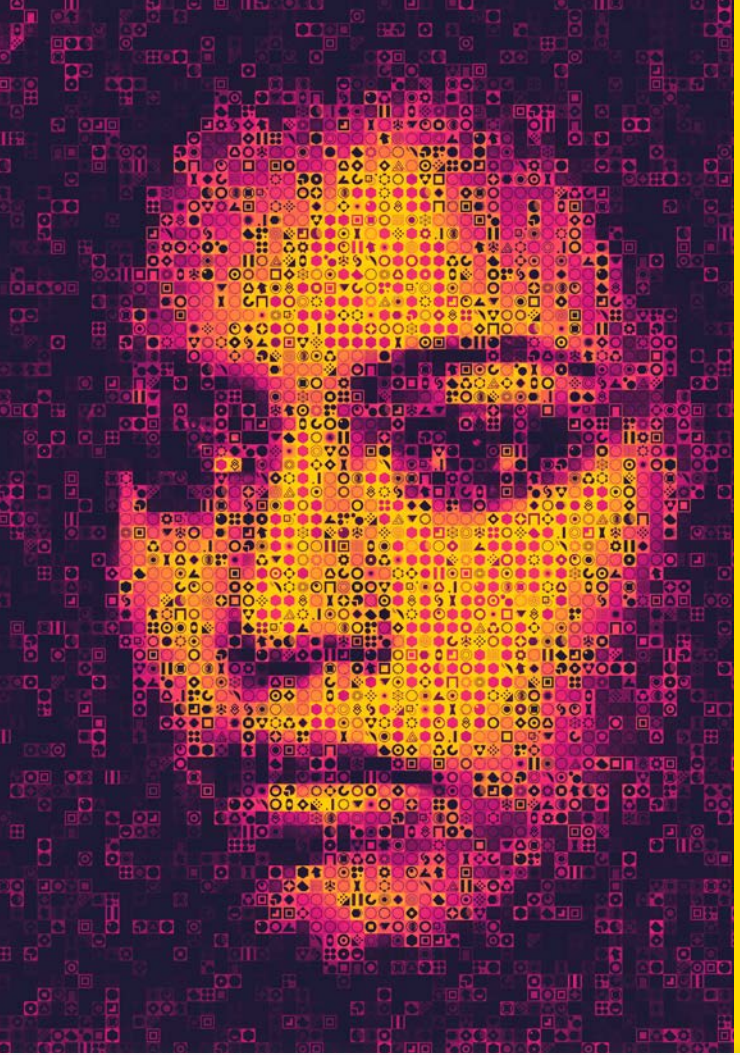
Alexis STEINER, *Chef de projets IA et Numérique de Grand E-Nov+*

MODE D'EMPLOI DU GUIDE

- Partie 1 : elle pose les définitions et le cadre dans lequel s'inscrit la démarche.

- Partie 2 : elle contient la méthodologie à proprement parler sous la forme de recommandations pratiques à mettre en œuvre en trois étapes :
  1. mise en place d'une gouvernance de l'IA au niveau de l'entreprise,
  2. évaluation du risque éthique du projet, identification des axes où porter l'effort à l'aide d'une grille de mesure de la sensibilité de son projet aux enjeux éthiques,
  3. application des recommandations en fonction des axes retenus, à chaque étape du cycle de vie.

- Partie 3 : elle vous propose des exemples de réalisations et des cas d'usage destinés à illustrer la mise en œuvre de la méthodologie.



PARTIE 1

CADRE ET REPÈRES

NOTES :

L'ENJEU

L'IA est porteuse d'un grand nombre de promesses : nous aider à relever nos grands défis sociétaux et environnementaux, optimiser le fonctionnement de nos entreprises, réduire la pénibilité de certaines activités, mieux répondre au client, diagnostiquer plus vite et mieux, nous assister dans nos vies quotidiennes, etc.

Pour autant, ces technologies, notamment les plus récentes dites de deep learning, suscitent le **questionnement du public**. L'enveloppe de mystère qui les entourent y est sans doute pour quelque chose. Mais les dérives de certains systèmes d'IA[1] et les dommages corporels provoqués par d'autres[2] ont exacerbé la méfiance des utilisateurs qui s'interrogent aujourd'hui sur la capacité des concepteurs de systèmes d'IA à maîtriser leurs créations.

1 - Outil de recrutement d'Amazon qui s'est avéré sexiste et chatbot Tai de Microsoft qui tenait des propos injurieux

2 - Voiture autonome Uber responsable d'un accident mortel

L'ENJEU

Pour éviter que ces réticences ne se transforment en un rejet systématique et massif, oblitérant du même coup les potentiels bienfaits de ces technologies, il devient urgent de penser ces dernières différemment.

Il faut désormais vouloir les systèmes d'IA non seulement performants mais aussi éthiques, c'est-à-dire conçus et pilotés de telle manière que leurs effets et leur emploi ne portent atteinte ni à la dignité ni à l'intégrité de l'être humain. Il faut également qu'ils respectent les valeurs éthiques fondamentales, telles que la préservation de la vie privée, l'équité dans les traitements ou la liberté d'agir et de décider.

Les professionnels du numérique, dont les métiers consistent à **concevoir, développer, exploiter et décommissionner des solutions d'IA**, sont concernés au premier chef par l'enjeu.

C'est pourquoi Numeum et ses partenaires vous proposent ce **guide pratique** détaillant une méthodologie pour créer des systèmes d'IA conformes aux valeurs éthiques et en mesure de répondre aux attentes de la société. C'est, en effet, à ce prix que pourra **s'instaurer, voire se restaurer, la confiance envers ces technologies**, condition indispensable à leur généralisation.

IA ÉTHIQUE : DE QUOI PARLE-T-ON ?

« La législation ne suit pas toujours le rythme des évolutions technologiques, ne correspond parfois pas à des normes éthiques ou peut simplement s'avérer inadaptée face à certaines questions. Pour être dignes de confiance, les systèmes d'IA devraient donc également être éthiques, en veillant à l'alignement sur les normes éthiques. »
Lignes directrices en matière d'éthique pour une IA de confiance, GEHN IA.

Un système d'IA peut se qualifier d'éthique s'il est en mesure de préserver, tout au long de son cycle de vie, les droits humains fondamentaux : droits à la dignité, à l'intégrité mentale et physique, à la liberté, à l'autonomie, à l'équité de traitement, à l'intimité et à la vie privée.

Ces droits sont en principe protégés par les lois et règlementations françaises et européennes en vigueur, comme le RGPD, qui régit la protection des données personnelles. Une IA, qu'elle se déclare éthique ou non, doit bien évidemment se soumettre à la législation sous peine d'être illicite et de ne pouvoir accéder au marché. Il faut cependant aller plus loin. La technologie progresse plus rapidement que le droit. Or, la nature des systèmes d'IA actuels, notamment ceux faisant appel aux technologies de deep learning, de même que leurs fortes incidences sur nos vies font émerger une concentration élevée de risques du point de vue de l'éthique : discrimination, déshumanisation des relations sociales, opacité des processus de déduction, utilisation abusive des données personnelles, erreurs provoquées par des cyberattaques, etc.

Ces risques sont à prendre en considération d'autant plus sérieusement que des arbitrages surviendront immanquablement pour résoudre les contradictions entre objectifs de performance et principes éthiques. La situation appelle donc des dispositions et précautions spécifiques pour éviter toute dérive volontaire ou involontaire.

Car, au bout du compte, l'Homme reste le seul responsable des systèmes qu'il crée.

IA ÉTHIQUE : DE QUOI PARLE-T-ON ?

« Tout personne impliquée dans la création d'une IA, à quelque étape que ce soit, est responsable de considérer l'impact du système sur le monde. »
Everyday ethics for artificial intelligence, IBM.

« En apprentissage par renforcement, l'évaluation ne devrait pas uniquement porter sur la capacité de l'IA ou du robot à atteindre l'objectif fixé, mais aussi sur sa capacité à produire le résultat attendu en se conformant aux valeurs humaines. »
The ethics of code, Sage.

En analysant différents textes *(voir bibliographie p.13)*, Numeum s'est attaché à caractériser l'IA éthique par un ensemble de qualités qu'elle doit rassembler. Une IA peut ainsi être qualifiée d'éthique si elle est :

1. **Respectueuse** des données personnelles.
2. **Équitable :** elle s'attache à ne pas créer ou reproduire de la discrimination et elle favorise l'inclusion.
3. **Transparente** dans son fonctionnement : ses conclusions peuvent s'expliquer et elle est auditable.
4. **Loyale** dans sa relation avec les êtres humains : elle ne fait que ce qu'on attend d'elle et elle se révèle à l'utilisateur.
5. **Maîtrisée :** elle reste sous le contrôle de l'humain.
6. **Sûre :** elle est sécurisée et robuste face aux cyberattaques.

Chacune de ces qualités reflètent un certain nombre d'exigences. Ce guide méthodologique, à vocation opérationnelle, décrit ces exigences et propose l'état de l'art des bonnes pratiques pour les remplir. Il appartiendra ensuite à chaque entreprise d'apprécier le risque de son système d'IA sur le plan éthique et de décider du degré d'application des mesures préconisées en fonction de ce risque et de la finalité du système.

Nota bene : *Une IA peut, par ailleurs, contribuer de manière positive à la société et au bien-être des personnes à travers sa finalité (AI for good). Nous n'abordons pas ce sujet dans le document dont la vocation est de recenser les mesures pratiques à mettre en œuvre pour créer et déployer une IA éthique by design, indépendamment de ce pour quoi elle a été conçue.*

BIBLIOGRAPHIE

Pour cadrer sa démarche et définir les axes de travail, Numeum s'est appuyé sur un corpus de chartes éthiques et de publications de référence existantes :

- Algorithmes, contrôle des biais SVP, livre blanc rédigé par l'institut Montaigne (mars 2020)
- Déclaration de Montréal pour un développement responsable de l'intelligence artificielle, rédigée par un collectif scientifique pluridisciplinaire et interuniversitaire de Montréal dans une démarche de co-construction et de concertation citoyenne (2018)
- Ethically aligned design (2017) et Establishing standards for ethical technology P70xx (2018), deux communications issues de l'IEEE
- Everyday Ethics for Artificial Intelligence, guide proposé par IBM (2019)
- Lignes directrices en matière d'éthique pour une IA digne de confiance (avril 2019) et Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment (juillet 2020), deux textes rédigés par le Groupe d'experts indépendants de haut niveau sur l'IA constitué par la Commission européenne
- IA responsable : principes, approches et mise en action, rédigée par Microsoft (2018)
- Renforcer la confiance dans l'intelligence artificielle axée sur le facteur humain, communication de la Commission au parlement européen, au conseil, au comité économique et social européen et au comité des régions (Com 2019 168, avril 2019), qui reprend les Lignes directrices citées ci-dessus
- Responsible AI: A Global Policy Framework, cadre proposé par l'organisation Itechlaw (2017)
- Rome Call for AI Ethics, l'appel du Vatican (février 2020)

- The Ethics of code. Developing AI for business with five core principles, charte rédigée par l'éditeur Sage (2017)
- Un engagement collectif pour un usage responsable de l'IA, charte rédigée par Impact AI
- Une approche européenne axée sur l'excellence et la confiance, livre blanc produit par la Commission européenne (COM 2020 65) (février 2020)
- Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), premier cadre juridique sur l'IA proposé par la Commission européenne (avril 2021)

PARTIE 2

RECOMMANDATIONS PRATIQUES

NOTES :

UNE GOUVERNANCE DE L'IA AU PLUS HAUT NIVEAU DE L'ENTREPRISE

L'enjeu lié au risque éthique d'un système d'IA est trop élevé pour reposer sur les seules têtes des data scientists et autres spécialistes techniques de l'entreprise. Le sujet est souvent d'une grande complexité. Il requiert différentes mises en perspective, notamment dès lors qu'on s'aventure dans une analyse d'impact. Il peut nécessiter des décisions de haut niveau qui réclament de réunir des compétences autres que celles présentes dans une équipe projet. Il peut, en outre, faire porter un risque réputationnel ou juridique sur l'entreprise.

C'est pourquoi la première recommandation de ce guide est la mise en place, au plus haut niveau, d'une gouvernance spécifique visant à définir et faire appliquer la politique de l'entreprise en matière d'IA. Reflétant les valeurs éthiques de l'entreprise et conforme à la règlementation, cette politique fixera les principes sur lesquels s'appuyer. Elle posera le cadre dans lequel les projets d'IA seront développés, c'est-à-dire un ensemble de processus et de méthodes de gestion de projet standardisés visant à sécuriser le traitement des questions éthiques.

Le système de gouvernance ainsi créé se verra confier la charge de définir une méthode d'évaluation des risques éthiques et de s'assurer de la réalisation des analyses d'impact requises. En fonction des projets, son rôle consistera aussi à arbitrer les tensions possibles (entre performance et enjeux éthiques et de sécurité), établir les seuils et autres limites, préciser les critères d'équité et d'explicabilité auxquels se référer et instruire les cas critiques.

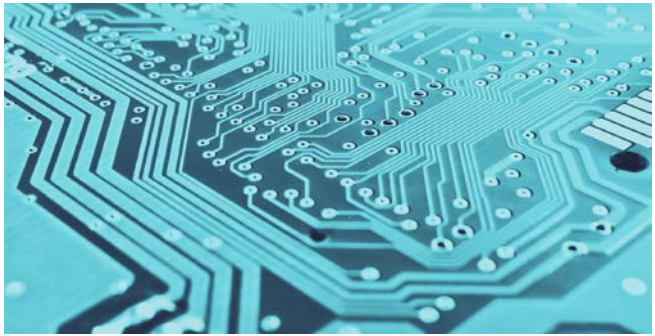
L'organisation et le périmètre de cette gouvernance varieront d'une entreprise à l'autre. Dans certains cas, les questions de conformité RGPD et NIS (Network and Information System Security) pourraient aussi lui incomber.

→ On peut se référer au Guide IA digne de confiance publié par Impact AI pour mettre en place la gouvernance IA au sein de son entreprise (chapitre 3) et pour l'auto-évaluer (chapitre 4).
→ Également, pour évaluer sa gouvernance IA : le questionnaire sur la gouvernance des algorithmes d'IA dans le secteur financier, de l'autorité de contrôle prudentiel et de résolution (ACPR).

UNE GOUVERNANCE DE L'IA AU PLUS HAUT NIVEAU DE L'ENTREPRISE

Quelques autres tâches auxquelles la gouvernance devra se consacrer :

- sensibiliser les équipes projet – les spécialistes de la data (data scientists, data analysts, etc.) et les métiers – aux enjeux éthiques de l'IA ainsi qu'aux questions règlementaires et de sûreté,
- pour chaque projet, établir le principe d'un responsable des traitements, point de contact de l'entreprise sur ces questions, vis-à-vis de l'intérieur et de l'extérieur,
- mettre en place un processus d'escalade si un risque éthique intervient ; cette mesure requiert l'instauration d'un climat de confiance au sein de l'entreprise et dans la relation avec les partenaires permettant à tout acteur de la chaîne de lancer une alerte.



→ **Un exemple :** chez Microsoft, un canal de communication dédié à la remontée d'alertes vers le comité éthique local permet à quiconque percevant un risque éthique sur un système d'IA en cours de développement de se manifester (par exemple, pour signaler que les conclusions du système d'IA concerné pourraient conduire à refuser un service substantiel, porter un préjudice ou violer des droits de la personne).

ÉVALUER LES RISQUES ÉTHIQUES DU PROJET

Créer une IA éthique by design, c'est avant tout concevoir une IA dont le risque de conséquences dangereuses ou préjudiciables sur le plan éthique est minimisé, voire annihilé. Pour chaque projet d'IA, une évaluation du risque éthique s'impose. L'analyse se déclinera selon les 6 qualités évoquées page 12 et leurs exigences associées. Elle portera autant sur les risques induits par le cas d'usage que sur ceux inhérents à la technologie et ceux liés au contexte du projet.
Elle appréciera le risque par rapport à sa vraisemblance et à sa gravité et le mettra en regard de la finalité du cas d'usage. L'étude devra également tenir compte des risques plus généraux de l'entreprise, liés à son secteur, son activité et son marché.

Les grandes typologies de risques à analyser :

► Concernant **le cas d'usage**
Le risque majeur est celui que ferait porter le résultat d'un système d'IA sur les individus, la société, l'environnement :

- risque d'un résultat conduisant à refuser ou limiter l'accès d'une personne à un droit fondamental (refus d'accès à un service essentiel par exemple) ;
- risque de créer une addiction, un enfermement de la personne ;
- risque de survenance de cas non prévus et potentiellement non maîtrisés pouvant porter atteinte à l'intégrité et/ou à la dignité humaine ou impacter l'environnement (ce risque peut apparaître sur les systèmes bénéficiant d'une certaine autonomie) ;
- risque de discrimination…

► Concernant **le système**
Les principaux risques sont :

- risque de non-conformité aux règlements de protection des données en vigueur (le RGPD, en Europe - certaines données peuvent s'avérer à caractère personnel au sens de la CNIL, alors même qu'elles n'apparaissent pas comme directement identifiantes, comme une adresse IP, par exemple),
- risque de cyberattaques entraînant la divulgation d'informations à caractère personnel et/ou un détournement de finalité,
- risque d'opacité du chemin de décision du système d'IA,
- risque lié à l'absence de traçabilité qui empêcherait la correction du système et/ou la recherche de responsabilités…

► Concernant le **contexte du projet**
Le risque principal est celui lié à un défaut de gouvernance IA dans l'entreprise :

- risque de méconnaissance des enjeux par les équipes et/ou l'entreprise, risque d'absence de processus de correction et de réparation,
- risque d'absence de référents…

ÉVALUER LES RISQUES ÉTHIQUES DU PROJET

Les résultats de l'étude serviront à orienter les réflexions et à décider des mesures à prendre pour réduire le risque. Ils permettront notamment de définir les seuils au-delà desquels une intervention humaine est nécessaire.
→ Un outil très complet d'analyse de risques : Ethics & Algorithms Toolkit. Réalisé par un collectif américain, il cible l'usage d'IA dans le secteur public.

La **matrice de sensibilité éthique** présentée dans ce guide fournit une grille de lecture des recommandations listées dans les pages suivantes. Elle vise à aider les concepteurs à orienter leurs efforts en fonction de la sensibilité de leurs systèmes d'IA aux enjeux éthiques.

→ Quelques outils et questionnaires complémentaires pour mesurer concrètement l'impact éthique de son système d'IA :

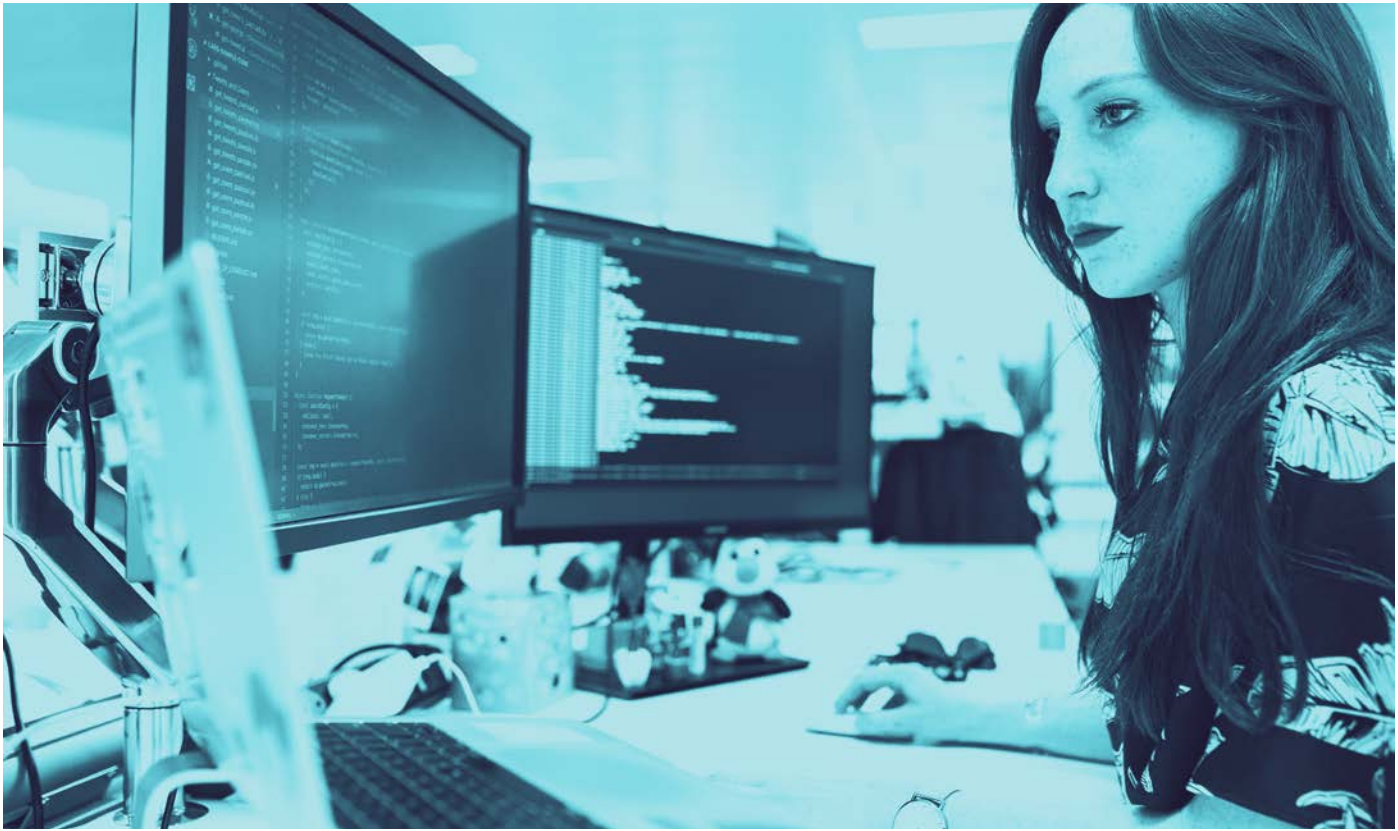
- Ethics Guide lines de la MAIF, intégré au projet Melusine de tri automatique d'emails.
- Le référentiel d'évaluation de la maturité d'une organisation, de Substra Foundation.
- Assessment List for Trustworthy Artificial Intelligence (Altai), le questionnaire du Groupe d'experts indépendants de haut niveau sur l'IA constitué par la Commission européenne.
- The Box de l'AI Ethics Lab qui permet de jauger selon différents critères les forces et faiblesses de son système sur le plan éthique (à lire en complément l'article *Operationalizing AI ethics principles*).
- Responsible AI Toolkit, un outil d'évaluation proposé par PWC.
- Artificial intelligence impact assessment, guide très complet de la plateforme indépendante néerlandaise ECP.

MATRICE DE SENSIBILITÉ ÉTHIQUE

Sujets éthiques à considérer spécifiquement →			Respect données personnelles		Équité			Transparence		Loyauté		Maîtrise	Sûreté
↓ Finalité et cadre de mise en œuvre du système		S'applique au projet	Confidentialité des données à caractère personnel	Usage encadré et mesuré des données à caractère personnel	Prévention contre les risques de discrimination	Diversité de l'équipe projet	Accessibilité de la solution	Explicabilité du modèle et des résultats	Traçabilité des données et des processus	Fiabilité des résultats	Dévoilement de l'IA	Fonctionnement sous contrôle humain	Robustesse et réslience de la solution
Le besoin métier	Le système automatise une décision, ou aide à prendre une décision, qui concerne des personnes physiques	Oui / Non			⊗			⊗	⊗	⊗		⊗	⊗
	Le système automatise l'exécution de tâches pour l'utilisateur	Oui / Non						⊗	⊗	⊗	⊗	⊗	⊗
	Le système est voué à se déployer à très large échelle - cf. interne à l'organisation vs (très) grand public	Oui / Non							⊗		⊗	⊗	⊗
	Le système est voué à se déployer sur un nouveau marché	Oui / Non			⊗				⊗		⊗	⊗	⊗
	Le système interagit directement avec l'utilisateur final	Oui / Non					⊗	⊗		⊗		⊗	⊗

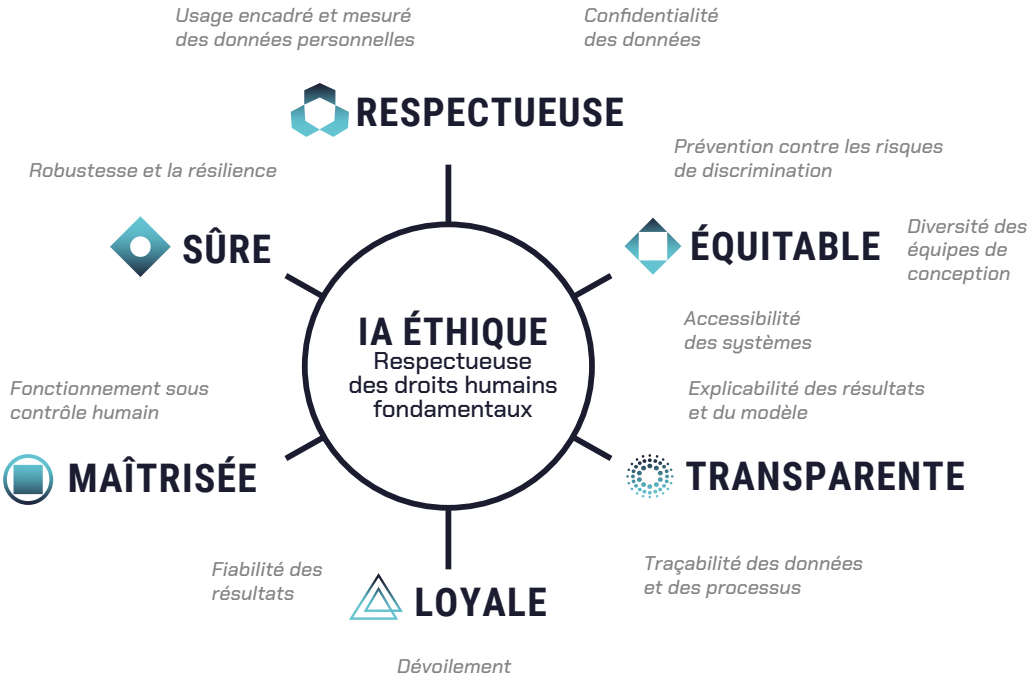
Sujets éthiques à considérer spécifiquement →		Respect données personnelles		Équité			Transparence		Loyauté		Maîtrise	Sûreté
↓ Finalité et cadre de mise en œuvre du système	S'applique au projet	Confidentialité des données à caractère personnel	Usage encadré et mesuré des données à caractère personnel	Prévention contre les risques de discrimination	Diversité de l'équipe projet	Accessibilité de la solution	Explicabilité du modèle et des résultats	Traçabilité des données et des processus	Fiabilité des résultats	Dévoilement de l'IA	Fonctionnement sous contrôle humain	Robustesse et réslience de la solution
La solution technique d'IA												
Le système est embarqué dans une solution plus large	Oui / Non							✕		✕		✕
Le système requiert un volume important de données pour s'entraîner	Oui / Non											✕
Le système requiert l'utilisation de données sensibles et/ou à caractère personnel pour s'entraîner	Oui / Non	✕	✕					✕	✕			✕
Le système requiert des jeux d'apprentissage provenant de bases de données publiques	Oui / Non			✕				✕				✕
Le système fait appel à une seule source de données pour construire son jeu d'apprentissage	Oui / Non											✕
Le jeu d'apprentissage est construit à partir de différentes bases de données hétérogènes (en termes de qualité, quantité, etc.)	Oui / Non			✕				✕				
Le système met en œuvre des technologies par nature non explicables (ou est susceptible de le faire)	Oui / Non						✕	✕				
Le système mobilise des briques technologiques « sur étagère »	Oui / Non											
Le système traite des données sensibles - cf. données à caractère personnel, données confidentielles, etc.	Oui / Non	✕	✕					✕	✕		✕	✕
Le système apprend en continu	Oui / Non						✕	✕		✕	✕	✕

Sujets éthiques à considérer spécifiquement →			Respect données personnelles		Équité			Transparence		Loyauté		Maîtrise	Sûreté
↓ Finalité et cadre de mise en œuvre du système		S'applique au projet	Confidentialité des données à caractère personnel	Usage encadré et mesuré des données à caractère personnel	Prévention contre les risques de discrimination	Diversité de l'équipe projet	Accessibilité de la solution	Explicabilité du modèle et des résultats	Traçabilité des données et des processus	Fiabilité des résultats	Dévoilement de l'IA	Fonctionnement sous contrôle humain	Robustesse et résilience de la solution
La gouvernance projet	L'équipe projet peut se référer à une instance dans l'entreprise en charge des sujets d'éthique et d'IA	Oui / Non	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	L'équipe projet peut se référer à des règles de gouvernance des projets d'IA	Oui / Non	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	L'équipe projet présente un défaut de diversité (genre, origine, culture, métier…)	Oui / Non			⊗	⊗	⊗						
	L'équipe projet a été sensibilisée aux enjeux de cybersécurité, et à ceux liés à l'IA particulier (cf. empoisonnement des données, attaques adversarielles, etc.)	Oui / Non		⊗								⊗	⊗
	L'équipe projet a été sensibilisée aux enjeux éthiques	Oui / Non	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	Certains acteurs de la chaîne de création du système sont des partenaires extérieurs	Oui / Non		⊗					⊗				⊗



LES BONNES PRATIQUES PAR QUALITÉ ÉTHIQUE

Cette partie décrit les bonnes pratiques qui conduisent au respect des exigences liées aux 6 qualités.
Elle détaille leur mise en œuvre à chaque étape du cycle de vie du système d'IA.



CYCLE DE VIE D'UNE SOLUTION IA

COMPRÉHENSION DU BESOIN

- Phase durant laquelle les membres de l'équipe projet, en particulier les data scientists, prennent connaissance du sujet et du cadre du projet auprès de leurs donneurs d'ordre ; c'est à ce moment que les questions éthiques sont abordées et que les différents risques, critères et seuils sont posés.

CONCEPTION

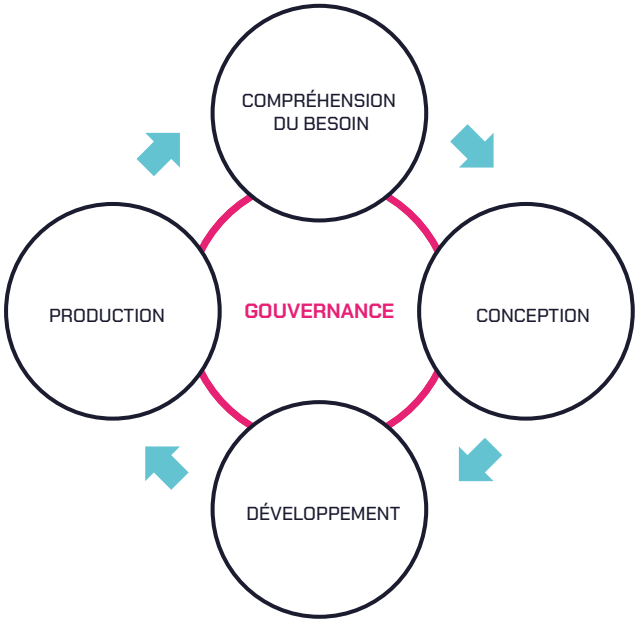
- Cycle itératif au cours duquel les data scientists construisent les différents jeux de données nécessaires à l'apprentissage et au test puis élaborent et testent le modèle.

DÉVELOPPEMENT

- Étape qui consiste à intégrer le modèle dans son environnement de production et à développer les modules complémentaires au modèle lui-même (les interfaces, par exemple) ; cette phase requiert en général des compétences informatiques.

PRODUCTION

- Le système d'IA est opérationnel ; il est monitoré et pourra faire l'objet de mises à jour ; en fin de vie, il sera décommissionné.



RESPECTUEUSE

« Les systèmes d'IA doivent garantir le respect de la vie privée et la protection des données tout au long du cycle de vie d'un système. Cela couvre les informations initialement fournies par l'utilisateur, ainsi que les informations générées au sujet de l'utilisateur au cours de ses interactions avec le système (par exemple, des résultats générés par le système d'IA pour des utilisateurs spécifiques, ou la manière dont les utilisateurs ont répondu à des recommandations spécifiques). » Lignes directrices en matière d'éthique pour une IA de confiance, GEHN IA.

EXEMPLES D'EXIGENCES

- La définition de la donnée personnelle diffère d'une région du monde à l'autre. Il en est de même des règlementations qui entourent la protection de ces données. La société conceptrice d'IA doit bien évidemment se conformer aux règlementations en vigueur dans les territoires où elle est présente. Au sein de l'Union européenne, le règlement à appliquer en la matière est le RGPD. Il repose sur 5 principes qui doivent guider la conception et la mise en œuvre des systèmes d'IA :
  - les données à caractère personnel sont utilisées dans un but précis et défini ;
  - le système ne collecte que les données à caractère personnel qui lui sont strictement nécessaires ;
  - la durée de conservation des données personnelles est raisonnable et le décommissionnement est prévu ;
  - des mécanismes de confidentialité et de sécurité protègent ces données ;
  - les personnes peuvent accéder à leurs données, les modifier, les supprimer et les transporter d'un traitement à un autre.
- Le système doit, en outre, respecter la vie privée et l'intimité des individus.
- Il ne doit pas, à l'insu des personnes, utiliser des données personnelles qu'il aura lui-même produites.

USAGE MESURÉ ET ENCADRÉ DES DONNÉES PERSONNELLES

En se nourrissant de quantités gigantesques de données, les systèmes d'IA qui fonctionnent par apprentissage automatique exacerbent les risques d'abus concernant les données personnelles. Ils défient les principes mêmes sur lesquels reposent les règlementations existantes de protection des données personnelles telles que le RGPD.

Pour le concepteur d'IA, le respect de ces principes soulève dès lors de multiples questions : quel équilibre entre données strictement nécessaires et performance ? Que signifie finalité dans une démarche d'expérimentation ? Quelle voie entre conserver les données pour les tracer et les supprimer ? Etc.

Pistes de solutions :

COMPRÉHENSION DU BESOIN

- Inciter les acteurs du projet à **minimiser l'emploi de données à caractère personnel** et, idéalement, à s'en passer.

- Si toutefois des données à caractère personnel sont nécessaires à quelque moment que ce soit du cycle de vie de la solution :

  - **Définir la base légale** sur laquelle va se fonder la collecte et le traitement de ces données à caractère personnel (contrainte règlementaire du RGPD).

    → Rappel des bases légales prévues par le RGPD (source : CNIL) :
    - le consentement : la personne a consenti au traitement de ses données ;
    - le contrat : le traitement est nécessaire à l'exécution ou à la préparation d'un contrat avec la personne concernée ;
    - l'obligation légale : le traitement est imposé par des textes légaux ;

- la mission d'intérêt public : le traitement est nécessaire à l'exécution d'une mission d'intérêt public ;
- l'intérêt légitime : le traitement est nécessaire à la poursuite d'intérêts légitimes de l'organisme qui traite les données ou d'un tiers, dans le strict respect des droits et intérêts des personnes dont les données sont traitées ;
- la sauvegarde des intérêts vitaux : le traitement est nécessaire à la sauvegarde des intérêts vitaux de la personne concernée, ou d'un tiers.

- **Effectuer un Privacy Impact Assessment (PIA)** pour évaluer l'impact des traitements sur ces données. Dans certains cas d'usage, par exemple s'il est question de données de santé, de la surveillance constante de personnes, d'outils de personnalisation et de ciblage, etc., l'exécution d'un PIA devient une exigence du RGPD. La CNIL met un outil à disposition à cet effet.

▶ Effectuer une analyse d'impact sur la vie privée de la **perte, de l'altération ou de la divulgation** des données manipulées (même à caractère non personnel).

CONCEPTION

- **D'une manière générale, maîtriser ses sources de données de tests et d'apprentissage** pour ne pas risquer de contrevenir à l'obligation de licéité du traitement des données (contrainte règlementaire, RGPD). Par exemple, éviter le web scrapping (qui consiste à collecter des data un peu partout sans vigilance).

- **Minimer, voire supprimer, l'usage de données à caractère personnel** en utilisant des **données synthétiques** le plus rapidement possible dans le processus de conception. Cette mesure s'avère notamment utile si un prestataire intervenant dans la conception de l'IA a besoin d'accéder aux données.

- Si l'utilisation de données à caractère personnel s'impose pour la conception et les tests :
  - les collecter et les traiter sur la base légale choisie (contrainte RGPD) ;
  - s'assurer de leur confidentialité (voir les différentes approches possibles plus loin) ;
  - mettre en place des dispositifs de contrôle d'accès (et surveiller le système en production via la journalisation des logs) pour empêcher le détournement de finalité par d'autres concepteurs (contrainte RGPD) ;
  - documenter les différentes mesures de mise en conformité du RGPD et notamment celles qui justifient l'emploi de données à caractère personnel pour la conception (finalité de l'application, contrainte d'accountability du RGPD).
    → Voir l'outil Datasheets for datasets

- **À noter que la conservation des données** qui ont servi à la conception du modèle (apprentissage et tests) peut s'avérer nécessaire (pour le traçage et pour l'accountability). Cela est possible **à condition toutefois**, en vertu du principe de durée de conservation des données inscrit dans le RGPD :
  - de justifier la conservation des données et d'informer sur sa durée,
  - de limiter la conservation à la stricte durée nécessaire à la réalisation de la finalité (ce qui implique de prévoir leur décommissionnement).
    → **Une idée :** créer des profils d'IA multiples correspondant à des sets de données d'entraînement différents pour les adapter aux exigences de l'utilisateur une fois en production (cette approche nécessite une étude de faisabilité techno-économique).

DÉVELOPPEMENT

- Prévoir la possibilité **d'arrêter la collecte** des données à tout moment si l'utilisateur le demande.
  → **Une idée :** bouton Stop Collect. Cela, quitte à proposer un fonctionnement limité ou moins performant du service à l'utilisateur, comme c'est le cas actuellement lorsqu'on refuse l'installation de cookies par un site web.
  - **Attention :** toutefois, si le refus de la collecte conduit à refuser ou restreindre l'accès à un service dit essentiel ou à un droit fondamental, prévoir une alternative permettant à l'utilisateur d'accéder à ce service.

- Développer des **interfaces claires** décrivant l'usage fait des données personnelles et permettant à l'utilisateur de **modifier/supprimer** à tout moment ses données.

PRODUCTION

- Garder le **contrôle sur les modèles déployés :**
  - s'assurer de manière régulière, en fonction de la sensibilité du système et des risques, que la finalité initiale reste conforme et que les dispositions prises en matière de respect des données personnelles restent respectées (exigence du RGPD d'effectuer ce contrôle une fois par an : contrôle de cohérence de finalité);
  - empêcher la réutilisation des modèles pour d'autres finalités ou par d'autres concepteurs (pour éviter le détournement de finalité (contrainte RGPD), par la mise en place de dispositifs de contrôle d'accès et par la surveillance du système en production via la journalisation des logs (voir chapitre Sûre/Robustesse et résilience).

- Mettre en place un **processus d'alerte auprès du responsable des traitements**, en cas de non-respect des exigences.

CONFIDENTIALITÉ DES DONNÉES PERSONNELLES

Le respect de la confidentialité des données requiert la mise en place de dispositifs de sécurité visant à empêcher l'intrusion illicite dans les bases de données (voir le chapitre Sûre/Robustesse et résilience). Il passe aussi par l'emploi de techniques empêchant de remonter jusqu'à l'individu à partir de ses données personnelles.

Plusieurs solutions existent. Prise individuellement, aucune n'est cependant totalement fiable. En outre, leur mise en œuvre tend à réduire les performances du système. Il faut également noter que, dans certains cas, la possibilité de réidentifier les personnes sur lesquelles le modèle a émis des conclusions peut s'avérer nécessaire (par exemple pour des algorithmes d'IA appliqués à la médecine personnalisée).

La bonne voie consistera souvent à combiner plusieurs approches, en attribuant plus de poids à l'une ou à l'autre en fonction de la finalité de l'application et des objectifs de performance.

Pistes de solutions :

CONCEPTION

- Appliquer, en les combinant, des méthodes visant à préserver la confidentialité des jeux de données.
  - **L'anonymisation :** il existe plusieurs techniques à choisir selon la finalité du cas d'usage et les caractéristiques des jeux de données. Si les techniques de base ne suffisent pas (si elles rendent possible la réidentification des personnes en recoupant avec des informations contenues dans d'autres bases, par exemple), des méthodes plus sophistiquées comme la confidentialité différentielle, qui consiste à générer du bruit autour des données collectées pour les noyer, peuvent être essayées. Une autre technique, dite d'anonymisation par avatar, employée par la société WeData, a été récemment approuvée par la CNIL.
  - **La pseudonymisation :** ces méthodes consistent à remplacer les données directement identifiantes d'un jeu de données (nom, prénom, etc.) par des données indirectement identifiantes (alias, numéro séquentiel, etc.). Il est également envisageable de collecter des données moins précises (tranche d'âge plutôt que l'âge, un code postal plutôt que l'adresse, etc.).

- **L'apprentissage distribué et/ou fédéré :** ces approches visent à réduire les points de centralisation des données, à garder ces dernières au plus près de l'entité qui les génère, et à ne plus les exposer.
  En apprentissage distribué, l'algorithme accède à une base de données qui reste localisée chez son propriétaire.
  En apprentissage fédéré, l'algorithme accède à un réseau distribué de bases de données (fonctionnement collaboratif).
  → Voir l'approche recommandée par Substra Foundation.
- **Le chiffrement des données :** c'est une approche courante dans les projets où des données confidentielles sont mutualisées. Les données de santé du Health Data Hub sont ainsi chiffrées.
  → Pour approfondir le sujet, voir les méthodes décrites par Substra Foundation.

▶ Protéger également la confidentialité des modèles (qui pourraient par inférence révéler les données utilisées dans l'apprentissage). Par exemple, en effectuant un apprentissage par distillation de modèles (qui a en plus le mérite de comprimer le modèle).

▶ Mesurer l'efficacité de l'anonymisation, par une analyse des risques de réidentification.
→ Par exemple à l'aide d'outils tels que celui d'ARX : Risk analysis - Data Anonymization Tool.

▶ Transférer à ses partenaires des sets protégés par les méthodes ci-dessus.

▶ Documenter les vulnérabilités et les techniques mises en œuvre pour y remédier.

POUR ALLER PLUS LOIN

- Audit requirements for personnal data processing activities involving AI : le guide méthodologique de l'autorité espagnole de protection des données pour évaluer la conformité d'un système d'IA au RGPD (2021).
- Guide RGPD du développeur de la CNIL.
- De la difficulté technique de l'anonymisation ou comment mal anonymiser ses données , un article publié sur Medium par Wavestone (2018).
- Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle que l'on peut trouver sur le site Éthique et intelligence artificielle de la CNIL (2017).

ÉQUITABLE

« Les ensembles de données utilisés par les systèmes d'IA (tant pour leur entraînement que pour leur exploitation) peuvent être biaisés par des partis pris historiques accidentels, des omissions et des modèles de gouvernance défectueux. La persistance de ces biais pourrait être source de discrimination et de préjudice (in)directs involontaires. »
Lignes directrices en matière d'éthique pour une IA digne de confiance, GEHN.

EXEMPLES D'EXIGENCES

- Le système doit fonctionner de manière impartiale. Il doit, en particulier, viser à ne pas renforcer ou créer de la discrimination du fait de biais introduits lors de l'entraînement ou dans l'algorithme.
- Il doit refléter la diversité de la population utilisatrice ou concernée.
- Il doit se conformer aux normes d'accessibilité les plus courantes et favoriser la diversité et l'inclusion.

PRÉVENTION CONTRE LES RISQUES DE DISCRIMINATION

Un système d'IA par apprentissage machine peut comporter des biais (par exemple, s'il a été entraîné avec un jeu de données lui-même biaisé). Le risque dans ce cas est d'aboutir à des résultats faux ou discriminatoires et donc préjudiciables. D'où l'exigence de recherche d'équité des systèmes. Pour autant, chercher à résoudre la question de manière totale et définitive peut relever de la quadrature du cercle. La notion est éminemment culturelle. Qui plus est, elle se présente sous différentes formes, parfois incompatibles entre elles. En fonction du cas d'usage, il sera donc nécessaire d'établir les critères d'équité que l'on souhaite obtenir en s'assurant qu'ils s'inscrivent dans le cadre éthique posé par l'entreprise, puis de mesurer les risques de biais et de tenter de s'approcher de la situation optimale.

Pistes de solutions :

COMPRÉHENSION DU BESOIN

- Si les conclusions du système d'IA concernent des personnes :
  - faire **une analyse des risques de discrimination et d'impact** (voir chapitre consacré à l'analyse des risques),
  - et, si un risque apparaît, examiner les paramètres et variables pouvant directement ou indirectement générer un risque de biais discriminatoire ou entraîner une dérive du modèle défini initialement ainsi qu'au fil de son utilisation.

- Le cas échéant, se rapprocher :
  - **de spécialistes des sciences humaines** (sociologues, anthropologues, etc.), qui apporteront leur expertise et aideront à identifier les risques de biais potentiels,
  - et/ou **de statisticiens** pour travailler sur la structure du jeu de données d'apprentissage et déjouer les pièges de la statistique.

CONCEPTION

▶ **Construction des jeux.**
→ **Une idée :** établir l'équité comme critère de succès (au même titre que la performance) pour orienter les travaux et garder la problématique à l'esprit lors de l'avancement.

- Dans tous les cas, **maîtriser ses jeux de données** d'apprentissage et de test : s'interroger sur leurs provenances et la manière dont ils ont été construits (connaître la source, la distribution du jeu de données, la façon dont les données ont été collectées, les transformations subies, etc.).
- Plusieurs pistes de construction :
  - créer des jeux très qualitatifs, en faisant appel à des experts du domaine considéré ;
  - employer des jeux d'entraînement existants et documentés ;
  - si les jeux ne sont pas représentatifs ou pas assez volumineux : utiliser les différentes techniques statistiques pour pallier le déficit (augmentation, rééchantillonnage, etc.) ou tester l'utilisation de réseaux génératifs (GAN) pour produire des données synthétiques ;
  - envisager de mutualiser ses données avec d'autres entreprises du même domaine, au niveau national et européen (voir l'exemple de Voice data).

▶ **Conception du modèle.**

- Intégrer des contraintes dans l'algorithme.

▶ **Mesurer et corriger.**

- Instrumenter le processus de tests pour faciliter sa mise en œuvre.
- Valider la représentativité du jeu de données : par exemple, en s'appuyant sur des référentiels existants (institutionnels, open data…).
- Tester l'algorithme selon les critères d'équité définis initialement.
  → Une approche défendue par l'Institut Montaigne dans son livre blanc Algorithmes, contrôle des biais SVP, et appelée équité active, consiste à utiliser des jeux contenant les variables protégées (l'approche nécessite une analyse d'impact auprès de la CNIL).
- S'assurer de ne pas avoir réintroduit de nouveaux biais en voulant corriger les premiers.
- Redresser l'algorithme d'apprentissage, le cas échéant.
  → **Une idée :** construire des algorithmes redresseurs de biais (en traitant biais par biais), à partir de GAN.

PRODUCTION

▶ Prévoir de surveiller les dérives en continu, avec des systèmes de supervision automatisés et des métriques appropriées (seuils, etc.).

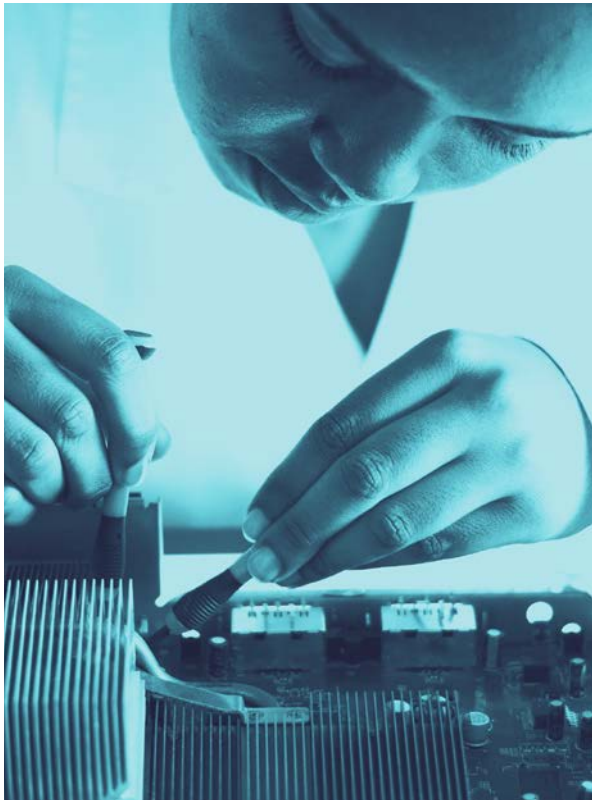
▶ Prévoir un point de contact permettant à l'utilisateur de remonter ses observations.

DIVERSITÉ DES ÉQUIPES DE CONCEPTION

L'exigence de diversité vise à aider les équipes projet à s'interroger sur les risques de biais de leurs produits dans les meilleures conditions et avec un état d'esprit le plus ouvert possible. Elle tend aussi à inciter les équipes à incarner les valeurs éthiques qu'elles défendent à travers leurs produits.

Pistes de solutions :

- Favoriser la diversité au sein des équipes (genre, culture et origine).
- Faire appel aux écoles inclusives.



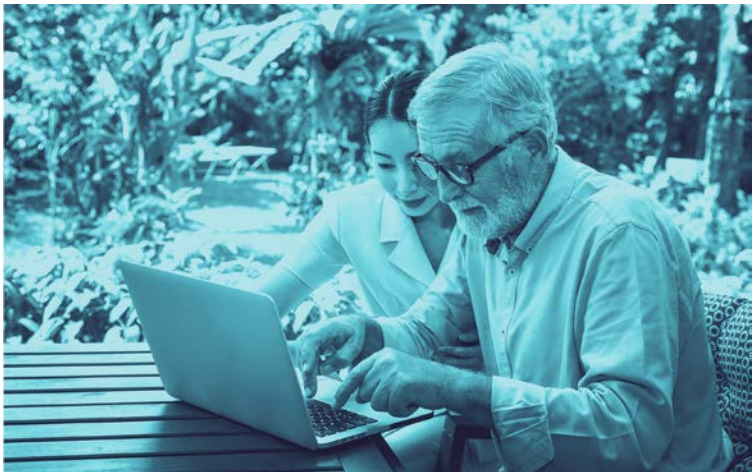
ACCESSIBILITÉ DES SYSTÈMES

L'IA présente un vrai potentiel pour faciliter l'accès au numérique de personnes souffrant de handicaps ou en difficulté par rapport au numérique. Au-delà des fonctions d'accessibilité règlementaires dont une interface graphique doit se prévaloir, doter les systèmes d'IA en interaction avec des humains de technologies de reconnaissance d'images, de text-to-speech ou d'analyse de textes, par exemple, favoriserait encore l'accessibilité de ces outils et l'inclusion.

Pistes de solutions :

CONCEPTION

- Se conformer aux **obligations règlementaires** en termes d'accessibilité des interfaces graphiques.
- S'appuyer sur le référentiel général d'amélioration de l'accessibilité.



POUR ALLER PLUS LOIN

- Algorithmes, biais, discrimination et équité, un livre blanc rédigé par Patrice Bertail, David Bounie, Stephan Clémençon et Patrick Waelbroeck de Télécoms ParisTech (février 2019).
- Algorithmes, contrôle des biais SVP, le livre blanc de l'institut Montaigne (mars 2020).
- Unfair biases in Machine Learning: what, why, where and how to obliterate them, un article sur l'origine des biais discriminatoires dans les algorithmes de machine learning et les techniques pour les résoudre, de Paul Irolla (MS Security, 2020).
- A tutorial on fairness in machine learning, un post technique de Ziyuan Zhong (Towards Data Science, 2018).

TRANSPARENTE

« Les individus doivent être en mesure de comprendre comment les systèmes IA prennent des décisions, surtout lorsque ces dernières ont un impact sur leur quotidien. »
Notre approche IA responsable, IA fiable, Microsoft.

EXEMPLES D'EXIGENCES

- Les causes et critères qui conduisent aux conclusions de l'IA doivent pouvoir être portés à la connaissance des utilisateurs et/ou des personnes affectés ou concernés par la conclusion ; ces explications doivent être intelligibles afin d'éclairer la décision finale ou de permettre d'agir sur les données d'entrée.
- Le processus de conception/ déploiement du système doit être documenté (description de la collecte et de l'étiquetage des données, de l'algorithme utilisé, du modèle économique, etc.) à des fins de vérification (audit) et d'amélioration/ correction de l'outil.
- Le processus de collecte, de conservation et d'usage des données doit être documenté (conformément au RGPD).
- Les différents arbitrages qui ont une incidence sur les exigences éthiques doivent être justifiés.

EXPLICABILITÉ DES RÉSULTATS

La capacité d'une IA à **rendre explicite l'impact d'une variable sur un résultat** et, plus généralement, la **compréhension des traitements** qu'effectue une IA, sont des exigences déterminantes pour l'acceptation des IA par la société. C'est parfois aussi une obligation de conformité et/ou de sécurité (la CNIL impose a minima d'informer sur les données utilisées pour arriver au résultat ; le RGPD impose l'explication si des données personnelles entrent en jeu dans le résultat). Malheureusement, les IA les plus performantes actuellement, dès lors que les volumes de données à traiter sont importants, se trouvent être les plus opaques... Des solutions pour résoudre cette question existent néanmoins. Elles peuvent être mises en œuvre en attendant que les nombreux travaux de recherche dans ce domaine portent leurs fruits.

Pistes de solutions :

COMPRÉHENSION DU BESOIN

- Estimer le besoin et le degré d'explicabilité requis, **en fonction du cas d'usage et de la finalité du produit.** La problématique se posera en effet de manière plus ou moins aigüe selon les cas d'usage. Ainsi, on accordera peut-être moins d'importance à l'explicabilité des résultats d'un service de recommandation de livres sur un site d'e-commerce qu'à un algorithme de scoring de prêt bancaire. Pour autant, un service d'e-commerce pourra souhaiter se distinguer par sa capacité à accompagner sa recommandation d'une explication dans un objectif éthique de lutte contre l'enfermement intellectuel.

- Si le besoin d'explicabilité s'impose, définir avec les métiers les **arbitrages à effectuer** entre précision des résultats et transparence. Dans beaucoup de cas, au moins au démarrage, l'emploi d'un algorithme naturellement explicable et moins précis peut suffire pour répondre au besoin.

- Envisager des systèmes hybrides, qui emboîteront des algorithmes de machine learning plus ou moins opaques dans des algorithmes à base de règles parfaitement interprétables.

- La question de l'explicabilité peut être reposée régulièrement tout au long de la conception au fur et à mesure de l'amélioration des résultats.

CONCEPTION

- S'il est nécessaire d'utiliser un algorithme complexe non naturellement explicable, s'orienter vers des **méthodes d'explicabilité a posteriori** (Lime, Shap...). Le choix de l'approche dépendra du **cas d'usage** et de **la cible de l'explication.**
  - Un client, un consommateur ou un utilisateur final souhaitera une explication correspondant à des variables précises (les siennes ou celles correspondant à un contexte précis). Le système devra fournir une explication locale (Lime ou Anchor, par exemple).
  - Un professionnel voudra comprendre le modèle dans son ensemble et attendra une explication globale en vue de l'améliorer (Shap, par exemple).
  - Le régulateur ou le superviseur voudra une preuve et aura donc besoin d'une explication globale (Shap, par exemple).

- **Tester** les différentes approches d'explicabilité pour en tirer les conclusions communes.

- **Documenter** les différentes approches testées et enregistrer les différents résultats pour en retenir les communs.

PRODUCTION

- Utiliser des **outils graphiques** qui permettent de visualiser les critères prépondérants.
  → Un outil tel que Shapash, conçu par la MAIF, met à la portée d'un non-spécialiste les résultats fournis par les outils d'explicabilité les plus courants (Lime, Shap...).

TRAÇABILITÉ DES PROCESSUS ET DES DONNÉES

Le traçage des jeux de données et des processus et méthodes de conception des différentes versions du modèle est la condition pour se trouver en mesure de vérifier l'absence de biais, le non-détournement de finalité et la fiabilité des systèmes. Cela signifie documenter tout ce qui a trait au système d'IA. Noter qu'il s'agit là, ni plus ni moins, que d'appliquer une démarche qualité, comme cela se pratique dans les autres domaines de l'informatique.

Pistes de solutions :

CONCEPTION ET DÉVELOPPEMENT

- **Documenter** :
  - la phase d'apprentissage : données utilisées (sources, transformation), paramètres et hyperparamètres, algorithme, versions, etc. :
    - on peut à ce sujet s'inspirer du model card de Google, sorte de carte d'identité du modèle,
    - ou de la méthode Datasheets for datasets suggérée par le collectif de chercheurs qui en est à l'origine ;
  - les différents arbitrages entre performances, explicabilité, confidentialité, sécurité ;
  - les acteurs impliqués dans la construction de l'IA et leurs rôles.

- Mettre en place **un référentiel** dans lequel seront stockées toutes les informations liées à l'IA.
  → **Une idée :** créer la généalogie du système à l'instar du modèle en cours d'élaboration de Substra Foundation.

PRODUCTION

- Continuer de documenter les données et comportements de l'IA en exploitation :
  - mettre en place un **système de journalisation** pour enregistrer les contextes d'obtention des résultats : l'algorithme, la version, le modèle, les paramètres et hyperparamètres, le jeu de données (sachant que cette opération peut rapidement se complexifier si le système apprend en continu…) ;
    → **Exemple d'outil** de versioning : DVC.org (Open-source Version Control System for Machine Learning Projects)
  - prévoir les **conditions de conservation** de ces informations (les logs peuvent, en effet, contenir des données personnelles ou sensibles*) : durée et finalité de conservation à préciser, mesures de sécurité à mettre en place, habilitations d'accès à délivrer, etc. (voir chapitre Respectueuse).

**Notons que, dans certains cas, l'application stricte du RGPD se heurte à de sérieuses difficultés opérationnelles : par exemple, si l'application est un système de cybersécurité qui enregistre des milliers d'adresses IP à la minute, lesquelles peuvent être considérées par la CNIL comme des données à caractère personnel.*

POUR ALLER PLUS LOIN

- Interpretable Machine Learning ou comment rendre les modèles boîtes noires explicables : un guide régulièrement mis à jour, par Christoph Molnar.
- Datasheets for datasets : un guide pour assurer le traçage de ses jeux de données, réalisé par un collectif de chercheurs de Google (en l'occurrence, Timnit Gebru), de Microsoft et de différentes universités (2020).

LOYALE

« Chaque personne doit savoir si elle interagit avec une machine. »
Rome Call for AI Ethics.

EXEMPLES D'EXIGENCES

- L'utilisateur doit comprendre sans l'ambiguïté qu'il interagit avec une machine.
- Le champ d'intervention, de même que les limites et capacités du système, doivent être portées à la connaissance de celui qui va le mettre en œuvre.
- L'utilisateur doit savoir si une IA est impliquée dans le résultat d'un calcul qui s'avère décisif pour lui ou qui pourrait orienter une décision qui le concerne ou qu'il doit prendre.
- Le système doit effectuer ce qu'on attend de lui, ni plus ni moins.

DÉVOILEMENT

La capacité du système à se dévoiler, c'est-à-dire à révéler ce qu'il est et ce qu'il fait, est une caractéristique clé pour instaurer la confiance des usagers et réduire les risques d'abus de faiblesse et d'addiction.
La problématique concerne les systèmes d'IA en interaction directe (chatbot, par exemple) ou indirecte (si le système est incorporé dans un outil) avec les humains. Il faut cependant noter qu'un système trop bavard s'exposera plus facilement aux attaques par inférence.

Pistes de solutions :

COMPRÉHENSION DU BESOIN

- Faire en sorte que toutes les parties prenantes du projet soient informées de l'enjeu lié au dévoilement du système afin qu'elles remontent les informations à communiquer et que soient définis les axes et canaux de communication vers l'utilisateur.

CONCEPTION

- Mettre en place un dispositif ou une méthode permettant d'informer l'utilisateur qu'il a affaire à une IA.
  → **Exemple :** dans le cas d'un chatbot, informer l'humain d'entrée de jeu qu'il interagit avec un robot ou concevoir une interface suffisamment explicite pour éviter l'ambigüité (par exemple, en évitant d'utiliser un avatar humain).

- D'une manière générale, fournir une synthèse lisible et compréhensible à l'utilisateur l'informant :
  - sur le fait qu'une IA est impliquée dans la solution utilisée ;
  - sur le type d'IA à l'œuvre, son périmètre d'intervention, ses objectifs ;
  - sur ses limites et la marge potentielle d'erreur ;
  - sur les conditions dans lesquelles elle doit être employée et les risques éventuels à l'utiliser ;
  - et, le cas échéant, sur l'usage que le système fait des données personnelles demandées (contraintes RGPD de finalité ; voir chapitre Usage encadré et mesuré des données personnelles).

FIABILITÉ DES RÉSULTATS

Ce sujet est le pendant du dévoilement selon lequel le système révèle ce qu'il est et ce qu'il fait. Ici, le système est jugé loyal et fiable s'il fait ce qu'il dit. La problématique figure parmi les plus complexes à traiter, car par nature, les IA qui fonctionnent par apprentissage automatique ne peuvent garantir à 100 % la reproductibilité de leurs résultats, a fortiori si elles apprennent en continu. Pour les systèmes complexes qui manipulent un très grand nombre de paramètres, on se heurte à la difficulté supplémentaire de ne pas pouvoir tester et donc valider l'ensemble des cas possibles.

Pistes de solutions :

COMPRÉHENSION DU BESOIN

- Établir des règles contraignantes et les limites à ne pas franchir au regard des résultats de l'étude de risques et d'impact réalisée en amont.

CONCEPTION

- Rédiger les spécifications détaillées du système et préciser les résultats attendus pour apporter au moins en partie la preuve de la fiabilité du modèle.

- Maîtriser le code :
  - faire des compromis entre simplicité et performance pour mieux maîtriser la reproductibilité des résultats ;
  - privilégier des codes documentés, en open source et de sources académiques (sachant qu'existe alors le risque de devenir la cible de cyberattaquants qui accèdent aussi au code).
    → **Une idée :** coupler des IA entre elles avec des méthodes de vote pour lisser les erreurs.

- Maîtriser les sets de données d'entraînement et de tests :
  - utiliser des sets de données d'entraînement documentés ;
  - auditer les sets et notamment évaluer la représentativité des données.

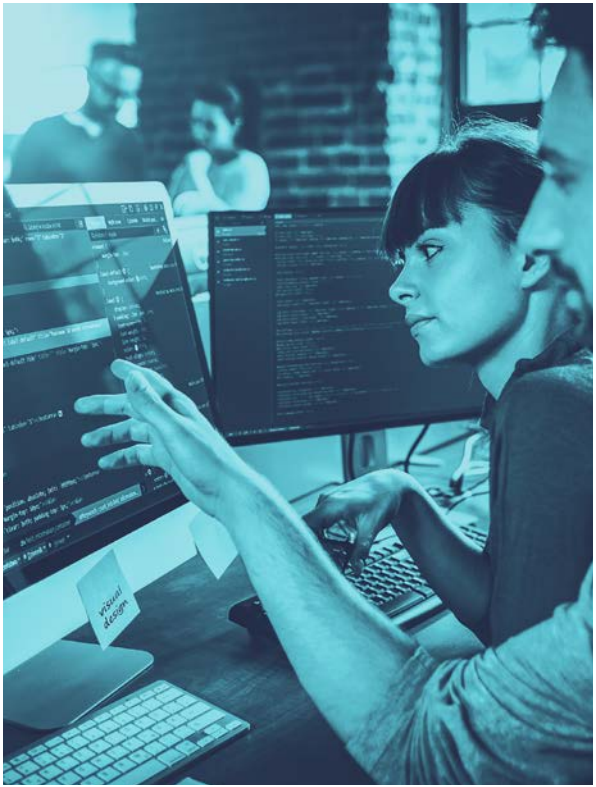
- Tracer (voir chapitre Traçabilité) :
  - tout documenter : source des data, traitement des data, modèle, algorithme, méthodes d'entrainement ;
  - créer une CI du modèle, qui va évoluer dans le temps.

DÉVELOPPEMENT

- Faire vérifier les résultats des systèmes d'IA par des experts métier pour garantir la fiabilité des résultats.
- Automatiser les tests pour les standardiser et faciliter leur exécution.
- Tester toutes les possibilités du système si cela est possible.
  - La capacité à mettre en œuvre cette mesure dépendra bien évidemment du niveau de complexité du cas d'usage. Envisager la mise en œuvre de GAN pour générer des données de tests, le cas échéant.

PRODUCTION

- Instaurer des protocoles de tests et de contrôle du comportement pendant toute la durée de vie du système pour s'assurer qu'aucune dérive ne se produit.
- Vérifier la viabilité des tests.



POUR ALLER PLUS LOIN

- Reproducible operations – commitment to the #4 principle, les pages du site de l'institut britannique The Institute for Ethical AI & Machine Learning consacrées à la fiabilité des systèmes de machine learning (le site, très complet, couvre l'ensemble des questions éthiques du machine learning).

- Le guide pour développer des chatbots conversationnels responsables de Microsoft.

MAÎTRISÉE

« Seuls des êtres humains peuvent être tenus responsables de décisions issues de recommandations faites par des SIA et des actions qui en découlent. »
Déclaration de Montréal.

« Dans tous les domaines où une décision qui affecte la vie, la qualité de la vie ou la réputation d'une personne doit être prise, la décision finale devrait revenir à un être humain et cette décision devrait être libre et éclairée. »
Déclaration de Montréal.

EXEMPLES D'EXIGENCES

- L'utilisateur d'un système d'IA prend connaissance des recommandations émises mais doit pouvoir rester décisionnaire et autonome dans ses choix individuels.
- Si la conclusion de l'IA doit conduire à une décision qui affecte une ou des personnes, alors la décision finale doit revenir à une personne.
- L'humain doit pouvoir prendre la décision de ne pas faire appel au système d'IA s'il juge que les conditions éthiques ou de sûreté ne sont pas remplies.
- Le système doit permettre à une personne d'émettre un recours ou de signaler une anomalie.

FONCTIONNEMENT SOUS CONTRÔLE HUMAIN

Risque de déshumanisation de nos sociétés et de perte d'autonomie de décision de l'individu, enjeu de responsabilité (l'humain reste le seul responsable des actions et décisions des IA)... Les raisons qui motivent ces exigences sont diverses. Du bouton « contact » à la restriction de l'automatisation du système, les mesures à prendre dépendront du cas d'usage.

Pistes de solutions :

CONCEPTION

- Si les résultats ont une conséquence sur l'humain, limiter au minimum l'automatisation du processus de décision. Le système doit rester un système d'aide à la décision.

DÉVELOPPEMENT

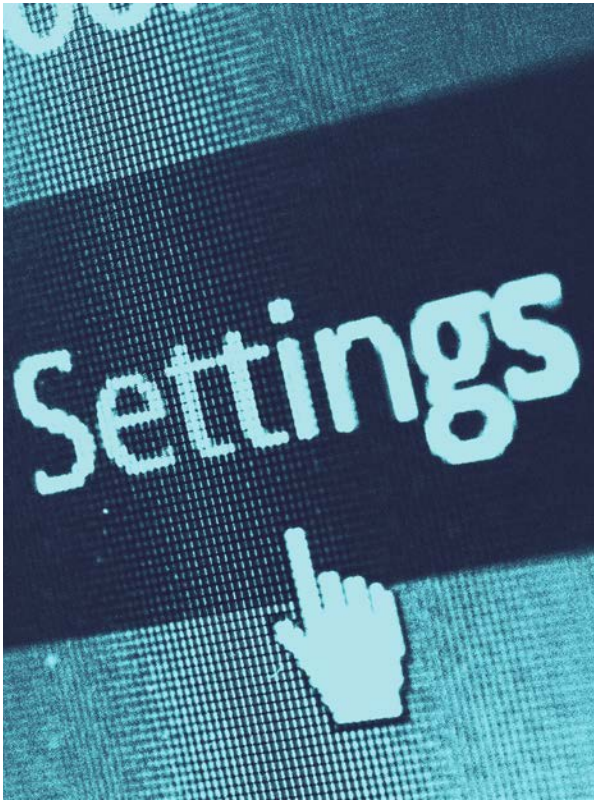
- Pour les systèmes orientés grand public, prévoir la possibilité d'ajuster les paramètres et données du modèle qui concernent l'utilisateur de manière interactive et temps réel, ou mettre en place un système permettant à l'utilisateur de valider ses paramètres lors de l'utilisation.

- Dans certains contextes grand public, notamment situés dans le domaine des services publics, informer de l'implication d'une IA dans le dispositif (en vertu de l'exigence de dévoilement) et prévoir la possibilité pour l'utilisateur d'avoir le recours de ne pas utiliser l'IA (contraintes du RGPD/art 22).

PRODUCTION

- Créer un bouton « Contact » permettant à l'utilisateur d'interagir, de faire remonter de l'information ou de déposer un recours et mettre en place le processus interne permettant de traiter les recours des utilisateurs ou les remontées d'informations.

  → **Une idée :** montrer à l'utilisateur l'ensemble des scores de l'IA et non pas seulement le résultat « définitif ». Cette approche suppose un apprentissage de l'utilisateur au fonctionnement de l'IA.



POUR ALLER PLUS LOIN

► Fully automated decision making AI systems : the right to human intervention and other safeguard, un guide méthodologique que l'on peut consulter sur le site Ai Auditing framework de l'autorité britannique de protection des données.

SÛRE

« Les systèmes d'IA et les environnements dans lesquels ils évoluent doivent être sûrs et sécurisés. Ils doivent être robustes sur le plan technique et il convient de veiller à ce qu'ils ne soient pas exposés à des utilisations malveillantes. »
Lignes directrices pour une IA digne de confiance, GEHN IA.

EXEMPLE D'EXIGENCE

- Le système doit comporter les mécanismes lui permettant :
  - de se prémunir contre les risques de malfaisance recensés lors de l'étude de risques et d'impact ;
  - de se protéger contre les cyberattaques (détournement de finalité, vol de données, etc.) ;
  - de bloquer et corriger les effets des éventuelles attaques et actions malfaisantes.

ROBUSTESSE ET RÉSILIENCE

Comme tout système informatique, les IA sont concernées par le cybercrime. Elles doivent donc bénéficier des mêmes règles, principes et dispositifs de protection que tout système d'information. Mais elles sont aussi l'objet de cybermenaces spécifiques :

- **Empoisonnement :** l'attaquant cherche à biaiser le comportement d'un modèle en modifiant les données d'apprentissage. Les systèmes apprenants en continu sont particulièrement exposés à ce type d'attaque.
- **Évasion :** l'attaquant modifie de manière imperceptible les données d'entrée de l'application pour leurrer le système et lui faire produire une décision différente de celle normalement attendue (ex : le panda). Les systèmes traitant des données d'entrée complexes comme les images sont particulièrement sensibles à ce type d'attaque.
- **Inférence :** l'attaquant assaille l'IA de requêtes pour comprendre son fonctionnement et saisir les paramètres clés, dans le but d'imiter le système.
  Les systèmes qui diffusent beaucoup d'informations s'exposent plus facilement à ce type d'attaque.

Comment se prémunir de ces menaces pour garantir un fonctionnement qui ne porte atteinte ni à l'intégrité des personnes ni à aucune des valeurs éthiques ?
La première mesure à prendre, indépendamment des projets, consiste à sensibiliser les data scientists et data analysts aux enjeux de cybersécurité. Contrairement aux informaticiens, cette population, principalement issue des filières mathématiques et statistiques, est naturellement moins préoccupée par ces questions (voir le chapitre Gouvernance).

Pistes de solutions :

CONCEPTION

- **Sécuriser l'apprentissage**, pour réduire l'exposition aux attaques par empoisonnement :
  - maîtriser les sources des données d'apprentissage ;
  - protéger les accès par des dispositifs de contrôle d'accès et d'habilitations ;
  - réduire au minimum la quantité de données, et notamment de données sensibles, nécessaires à l'apprentissage (usage de données synthétiques, par exemple) ;

- appliquer les techniques permettant de renforcer la confidentialité des données sensibles (techniques d'anonymisation par confidentialité différentielle, de pseudonymisation et d'apprentissage distribué et/ou fédéré - voir chapitre La confidentialité des données) ;
- surveiller en continu l'évolution de l'apprentissage et les évolutions de comportement des modèles ;
- mettre en place des garde-fous : par exemple, pour un chatbot, créer une liste noire de termes à bloquer, en entrée et en sortie ;
- appliquer des pratiques de type RONI (Reject on Negative Impact), qui consistent à rejeter les données qui font dériver le modèle.

► **Renforcer la robustesse des modèles** pour réduire leur sensibilité aux différentes attaques, notamment aux attaques par inférence et par évasion.

- Différentes techniques existent : distillation de modèles, apprentissage adversarial, bruitage des données d'entrée, etc. Elles seront à appliquer en fonction du niveau de précision que l'on veut atteindre.
- Réduire au minimum les informations qui pourraient révéler le fonctionnement du modèle (score, précision, etc.) pour réduire les risques d'attaques par inférence. Cette mesure s'appliquera par rapport au niveau de dévoilement que l'on estime indispensable.

DÉVELOPPEMENT

► **Tester :** mettre en place des processus (automatisés) et des outils comme la boîte à outils ART 360 d'IBM ou la démarche d'altération de code.

PRODUCTION

► **Surveiller :** mettre en place un processus de surveillance du comportement et des logs en temps réel. S'assurer notamment que les résultats ne dérivent pas, en faisant vérifier régulièrement les comportements et résultats par des experts métier.

► Créer un bouton contact permettant à un utilisateur d'alerter le responsable des traitements.
→ **Une idée :** créer une plateforme Bug bounty.

POUR ALLER PLUS LOIN

- Intelligence artificielle et cybersécurité, un livre blanc de Wavestone (2019).

PARTIE 3

CAS D'USAGE ET EXEMPLES

NOTES :

CAS D'USAGE E-COMMERCE

CONCEPTION ET MISE EN ŒUVRE D'UN ALGORITHME DE RECOMMANDATION POUR UN SITE VENDANT DES PRODUITS CULTURELS EN LIGNE.

Un client potentiel (prospect) se rend sur un site d'e-commerce vendant des produits culturels. Il consulte le catalogue de produits proposés pour y trouver un article correspondant à ce qu'il recherche. Le site lui recommande des produits durant sa navigation.

RISQUES ÉTHIQUES IDENTIFIÉS

Exemples de problématiques spécifiques à traiter pour l'entreprise d'e-commerce :

- **discriminations** liées au genre / aux convictions religieuses / au niveau de vie, etc. du client ;
- **enfermement** dans une bulle personnelle de contenus personnalisés ;
- **influence** par l'interface ;
- **utilisation de biais inconscients** du client pour influencer son achat (nudge).

Description du cas d'usage

Objectifs (vis-à-vis de l'utilisateur)	Enrichir l'expérience du client en personnalisant les contenus qui lui sont présentés pour répondre à son besoin, sans pour autant biaiser la sélection qui lui est soumise qui pourrait limiter son domaine de choix.
Finalités (pour l'entreprise qui recours à l'IA)	Augmenter ses ventes (maximiser le taux de conversion, optimiser le panier moyen, etc.).
Comment	En transformant le prospect en client en rendant effectif l'acte d'achat. Pour ce faire, il faut donc lui proposer le produit qui correspond le mieux à ce qu'il cherche en termes de produit lui-même, de prix, d'accès (collecte / livraison…), etc.
Actions mises en œuvre	Optimiser la présentation du produit - niveau de description du produit, photos, etc. Présenter des produits similaires et/ou complémentaires.
Solution retenue (type d'IA)	Modèle / Algorithme de recommandation.
Méthodes / Outils utilisés	Filtrage collaboratif (user based / item based) Content based (proximité par typologie de produits). Populaire (modèle d'asssociation - cf. association de produits dans l'acte d'achat).
Métiers concernés	Lors de la compréhension du besoin : marketing produits, distributeurs, producteurs produits, consommateurs. Lors de la conception : datascientists, dataengineer, designer. Lors du développement : developpeur, testeurs. Lors du déploiement : superviseur, testeurs.

Évaluation de la sensibilité de la solution aux questions éthiques (selon la matrice de sensibilité éthique)

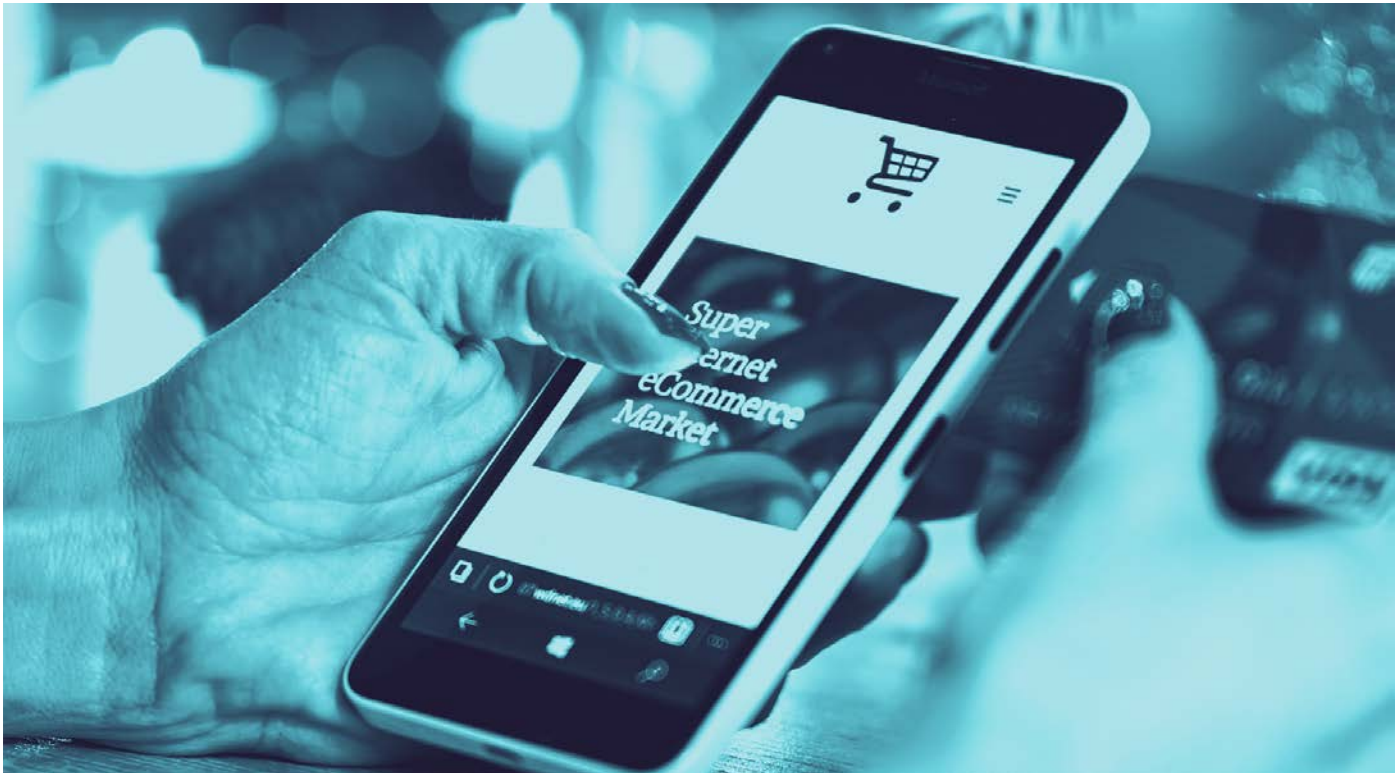
FINALITÉ ET CADRE DE MISE EN ŒUVRE DU SYSTÈME

		Applicable au projet ?
Le besoin métier	Le système automatise une décision, ou aide à prendre une décision, qui concerne des personnes physiques	Oui
	Le système automatise l'exécution de tâches pour l'utilisateur	Non
	Le système est voué à se déployer à très large échelle - cf. interne à l'organisation vs (très) grand public	Oui
	Le système est voué à se déployer sur un nouveau marché	Non
	Le système interagit directement avec l'utilisateur final	Oui
La solution technique d'IA	Le système est embarqué dans une solution plus large	Oui
	Le système requiert un volume important de données pour s'entraîner	Oui
	Le système requiert l'utilisation de données sensibles et/ou à caractère personnel pour s'entraîner	Non
	Le système requiert des jeux d'apprentissage provenant de bases de données publiques	Non
	Le système fait appel à une seule source de données pour construire son jeu d'apprentissage	Non
	Le jeu d'apprentissage est construit à partir de différentes bases de données hétérogènes (en termes de qualité, quantité, etc.)	Oui
	Le système met en œuvre des technologies par nature non explicables (ou est susceptible de le faire)	Non
	Le système mobilise des briques technologiques « sur étagère »	Oui
	Le système traite des données sensibles - cf. données à caractère personnel, données confidentielles, etc.	Non
	Le système apprend en continu	Oui
La gouvernance du projet	L'équipe projet peut se référer à une instance dans l'entreprise en charge des sujets d'éthique et d'IA	Oui
	L'équipe projet peut se référer à des règles de gouvernance des projets d'IA	Oui
	L'équipe projet présente un défaut de diversité (genre, origine, culture, métier…)	Oui
	L'équipe projet a été sensibilisée aux enjeux de cybersécurité, et à ceux liés à l'IA particulier - cf. empoisonnement des données, attaques adversarielles, etc.	Non
	L'équipe projet a été sensibilisée aux enjeux éthiques	Oui
	Certains acteurs de la chaîne de création du système sont des partenaires extérieurs	Oui

Sujets à traiter
(selon la matrice de sensibilité éthique)

Pistes de solutions identifiées
(pour chaque sujet à traiter)

SUJETS ÉTHIQUES À CONSIDÉRER SPÉCIFIQUEMENT		ACTION(S) / PHASE(S) DU PROJET / ACTEUR(S)
Respect	Confidentialité des données à caractère personnel	**Conception :** pseudonymisation des données personnelles, mise en place d'accès sécurisé aux données, liste des features en entrée des algorithmes, travail à partir d'identifiant client, non intégration de données dites «sensibles» au sens du RGPD
	Usage encadré et mesuré des données à caractère personnel	**Conception :** deux traitements de données distincts : l'apprentissage et l'usage du modèle
Equité	Prévention contre les risques de discrimination	**Développement :** validation des données pertinentes pour le fonctionnement, minimisation finale
	Diversité de l'équipe projet	**Conception :** introduction d'aléas pour de la diversité volontaire dans la recommandation **Développement :** diversité dans l'équipe de tests ou les rôles de tests
	Accessibilité de la solution	**Conception :** penser l'accessibilité à tous les publics dès le début
Transparence	Explicabilité du modèle et des résultats	**Conception :** besoin de transparence pour le testeur et pour le client **Déploiement :** proposer différents paramétrages à l'utilisateur : pertinence, dates…
	Traçabilité des données et des processus	**Conception et développement :** mise en place d'un data catalogue, de documentation projet (MCD, processus d'alimentation des données, etc.) pour assurer la traçabilité dans un outil de gouvernance
Loyauté	Fiabilité des résultats	**Conception :** définition de l'objectif et des métriques **Développement :** mesure et initiation du suivi **Déploiement :** suivi des performances, remontée d'alertes
	Dévoilement de l'IA	**Conception et développement :** prévoir un message d'alerte à destination de l'utilisateur final
Maîtrise	Fonctionnement sous contrôle humain	**Déploiement :** chaîne de supervision, suivi, boucle de feedback utilisateur via la relation client
Sûreté	Robustesse et résilience de la solution	**Conception :** définir des métriques **Déploiement :** mettre en place les processus de suivi et de traitement des alertes, backtesting régulier des modèles


Super
ernet
eCommerce
Market

EXEMPLE DE RÉALISATION

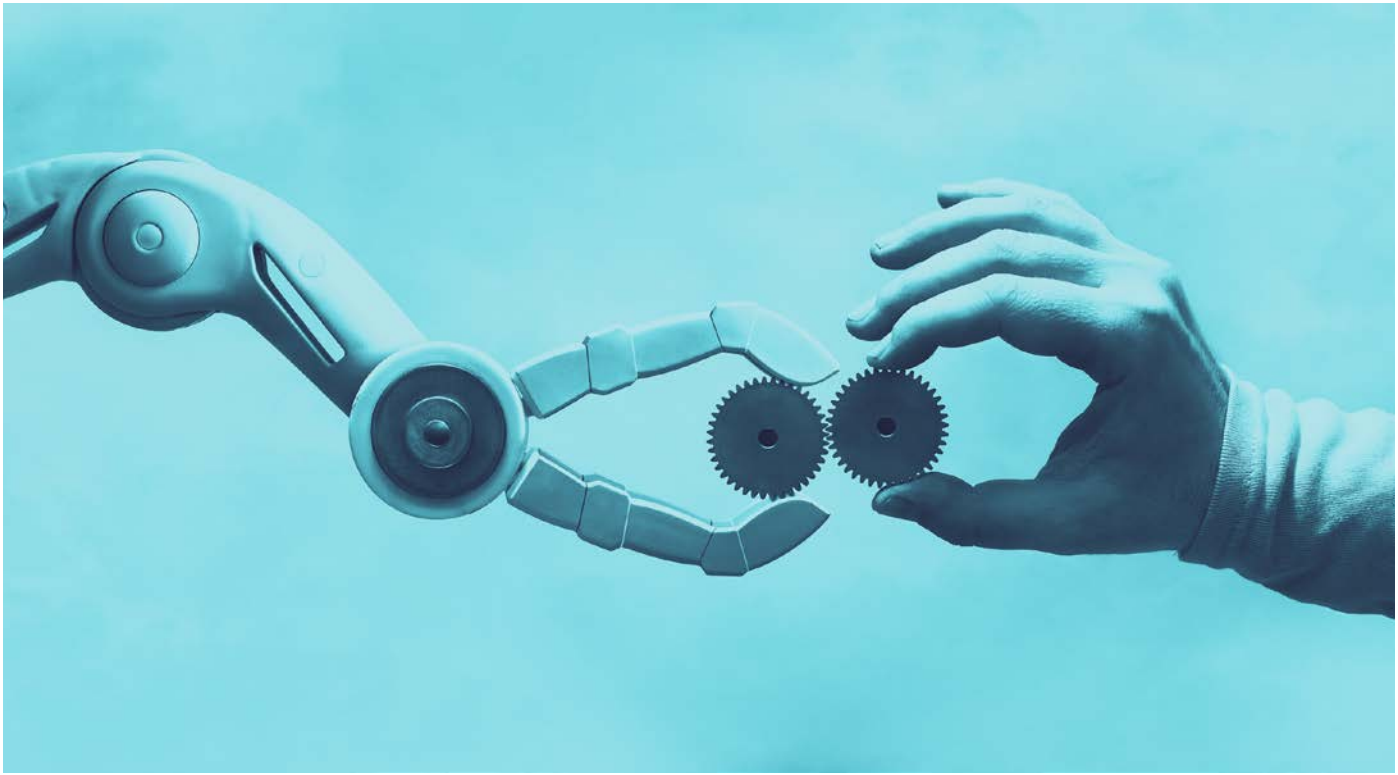
MISE EN ŒUVRE D'UN SYSTÈME DE TRAÇABILITÉ

- Certains scénarios de développement d'intelligences artificielles nécessitent des précautions particulières d'utilisation. Un exemple typique d'un tel cas relève des projets critiques pour lesquels l'IA agit en temps réel et sans validation humaine, entraînant un risque d'atteinte à la réputation du client ou de dommages matériels et humains dans le cas où le système viendrait à dériver.
  Un autre scénario qui mandate des précautions particulières se produit dans les cas où l'équipe chargée de l'intégration du modèle est distincte de l'équipe de R&D ayant créé le modèle en question, entraînant un défaut de communication potentiel : le risque est notamment accru dans le monde de l'Open source ou dans des projets de grande envergure comportant des unités opérationnelles séparées.

- Pour répondre à ces problématiques de cadrage strict concernant l'apprentissage et l'utilisation des modèles d'intelligence artificielle, nous utilisons ainsi une organisation standardisée sous forme de « model cards » proposée par les équipes de recherche de Google en 2019, dont l'objectif est de documenter, de manière succincte mais complète, un modèle d'IA. Une model card se présente sous la forme d'un résumé d'une vingtaine de lignes, à la manière du readme proposé avec certains logiciels, qui regroupe notamment les informations suivantes :
  - détails du modèle, et notamment : identité des développeurs, date de publication, paramètres du modèle, version, licence ;
  - cas d'usage prévus et conditions nominales d'utilisation ;
  - caractéristiques des jeux de données utilisés pour l'apprentissage ;
  - performance du modèle et jeux de données utilisés pour l'évaluation ;
  - considérations éthiques, mises en garde et recommandations.

- L'adoption de ce standard permet d'avoir un template commun de documentation de chaque intelligence artificielle, permettant une transmission maximale de l'information entre l'ingénierie et le développement pour assurer que les modèles sont utilisés correctement.Cela réduit par conséquent le risque de comportement imprévu des systèmes automatisés qui leur sont associés.

Mathis HAMMEL
Head of Cybersecurity R&D Sogeti (Groupe Capgemini)



Numeum et ses partenaires remercient chaleureusement tous les contributeurs aux travaux.

▶ COMITÉ DE PILOTAGE

- Céline BAYLE (SAGE)

- Bénédicte DE LINARES (BDL CONSEIL)

- Valentin HUEBER (NUMEUM)

- Katya LAINÉ (TALKR.AI BY KWALYS)

- Jean-Claude REMBERT-BAUDET (ASTEK)

▶ MENTORS ET ANIMATEURS

- Magali BARNOIN (TELECOM VALLEY)
- Benoît BOUFFARD (WAVESTONE)
- Marine BROGLI (DPO CONSULTING)
- David CORTES (AI-VIDENCE)
- Laurence DEVILLERS (SORBONNE UNIVERSITÉ/CNRS)
- Sébastien JARDIN (IBM FRANCE)
- Mouchira LABIDI (FREELANCE)
- Charlotte LISCHER (CATALIX)
- Alice LOUIS (CABINET DICÉ)
- Jean-Luc MAINGUY (SEENAPSYS)
- Fabrice MARQUE (ZEBRAVALLEY)
- Emmanuel NARS (DOCAPOSTE)
- Vincent PERRIN (IBM FRANCE)
- Françoise SOULIÉ (HUB FRANCE IA)
- Florence TRESSOLS (IBM FRANCE)
- Félicien VALLET (CNIL)

▶ POUVOIRS PUBLICS

- Renaud VEDEL (CSN-IA)
- Nicolas AMAR (CSN-IA)
- Martin BIERI (CNIL)

▶ CONTRIBUTEURS

- Sonia ABECASSIS (IBM FRANCE)
- Cindy ACCOLAS (GRAND ENOV +)
- Didier AÏT (OPTIM'EASE)
- Marianne ALLANIC (ALTHENAS)
- Aziz AMAL (ASTEK)
- Nadia ANGLESSY (NETSYSTEM SOLUTIONS)
- Nicolas Andréa ARZOTTO (LEADIN)
- Marion BALAC (ESAM)
- Franck BARDOL (DIAG)
- Renaud BAUVIN (CRITEO)
- Julie BEC (AIR FRANCE KLM GROUP)
- Jérôme BERANGER (ADELIAA)

- Gwenaëlle BODILIS (DPO SYSTEM)
- Marina BOECHAT (MYDATAMODELS)
- Eric BONIFACE (SUBSTRA FOUNDATION)
- Guillaume BUFFET (U CHANGE)
- Anne-Christine CARPENTIER (GFII)
- Pierre CHARARA (TESSI)
- Lucas CHARRON (SPORTINTECH)
- Edouard CHOPLAIN (C2IP)
- Tawhid CHTIOUI (AIVANCITY)
- Eugénie CLÉMENT (OCCITANIE DATA)
- Sophie COMPAGNON (CRITÉO)
- Jean-Baptiste CONAN (KEYRUS)
- Nathalie COSTA (YSANCE)
- Rébecca DADI (DPO CONSULTING)
- Guillaume DE LA ROCHE (RENAULT)
- Nathalie DELBECQ (RENAULT)
- Paul DESIGAUD (WAVESTONE)
- Alix FAUQUES DE JONQUIERES (ANITI)
- Sébastien FORET (GRAND ENOV +)
- Mickaël GADOUD (WAVESTONE)
- Mithuran GAJENDRAN (WAVESTONE)
- Nicolas GEORGEAULT (ASI)
- Guillaume GIMONNET (WAVESTONE)
- Emmanuel GOFFI (INSTITUT SAPIENS)
- Amélie HELIOU (CRITEO)
- Laëtitia KAMENI (ACCENTURE)
- François KLIEBER (BOUYGUES CONSTRUCTION)
- Djémila KOHIL (LPCE BIOBANK CÔTE D'AZUR)
- Bradreddine LADJEMI (ANKABOOT)
- Pascal LAINÉ (TALKR.AI BY KWALYS)
- Yanelle LARIBI (IMPACT AI)
- Yann LE BIANNIC (SAP)
- Fabrice LE GUEL (RITM)
- Frédéric LEBLAN (3DS OUTSCALE)
- Xavier LECLERC (DPMS)
- Bertrand LEJEUNE (CAP DIGITAL)
- Simon LEROY (KEYRUS)
- Clément LOMBARD (WAVESTONE)
- Daphné MARNAT (TWISTING)
- Laura MARTI (BOUYGUES CONSTRUCTION)
- Maud MARQUIS (MIO&CO)
- Didier MASCARELLI (KADLOG)
- Clément MAYER (SUBSTRA FOUNDATION)
- Igor MEKHOV (CONSORT NT)
- Stéphan MIR (WAVESTONE)
- Assia MOULOUDI (SAP)
- Claire NICODEME (SNCF)
- Bernard OURGHANLIAN (MICROSOFT)

- Pierre PARREND (EPITA)
- Alexandre PASCAULT (ASTEK)
- Stéphane PAULIN-HENRIKSSON (CNRS)
- Gaëlle PICARD-ABEZIS (DOCAPOSTE)
- Estelle PINCHEZON (HUMAN DESIGN GROUP)
- Gaëlle PINSON (HUB FRANCE IA)
- Marc PLATINI (GRAND ENOV +)
- Timothée RAYMOND (LINEDATA)
- Bernardo RESENDE (THALES SERVICES SAS)
- Bettina REVEYRON (IMPACT AI)
- Caroline RICHARD (NATIXIS)
- Laurent RISSER (ANITI)
- Céline RODAP (ECOLE 42)
- Roxana RUGINA (IMPACT AI)
- Laura SARRIOT (KILOUTOU)
- Céline SAVOY-LAMOTTE (TESSI)
- Anthéa SERAFIN (OCCITANIE DATA)
- Emilie SIRVENT-HIEN (ORANGE)
- Camille SOUILLART (HUB FRANCE IA)
- Thomas SOUVERAIN (DREAMQUARK)
- Alexis STEINER (GRAND ENOV +)
- Aurélia SZYMANSKI (LINEDATA)
- Lucien TANGHE (ASSURACTIS SARL)
- Eric TORDJEMAN (INRIA - INSTITUT DATAIA)
- Stéphanie TOUSSAINT (GRAND ENOV +)
- David TSANG-HIN-SUN (KEYRUS)
- Laura VELASCO AVISBAL (LABORATOIRES SERVIER)
- Eric VESSIER (ORACLE FRANCE)
- Richard VIDAL (ACCENTURE)
- Clément VIDON (SOCIÉTÉ CIVILE)
- Coline YVERGNIAUX (DEVOTEAM)

Initiative conduite par :

148, Bd. Haussmann - 75008 paris
01 44 30 49 70 - contact@numeum.fr
@numeum

Soutenue par :

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Premier ministre

3iA Côte d'Azur
Institut interdisciplinaire d'intelligence artificielle

42

aivancity
SCHOOL FOR TECHNOLOGY, BUSINESS & SOCIETY
PARIS-CACHAN

ANITI
Université Fédérale
Toulouse Institute

GRAND ENOV+

HUB FRANCE IA

IMPACT AI

INSTITUT DATAIA
Science des données, Intelligence & Société

Grand Est
ALSACE CHAMPAGNE-ARDENNE LORRAINE
L'Europe s'invente chez nous

Telecom valley
Animateur Azuréen Numérique

Conception graphique et réalisation : MAP ADVERTISING