

axionable

datacraft*

Retour d'expériences

Certification processus IA et labellisation IA responsable et de confiance



Avec vous aujourd'hui



Gwendal BIHAN

Président & CTO



José SANCHEZ

Deputy-CTO,
Senior Manager MLE

ACCÉLÉRER LA TRANSITION DURABLE GRACE A LA DATA ET L'IA



100% indépendant



50

COLLABORATEURS
IA / Tech / Climat



PARIS



MONTRÉAL

CLIENTS



INDUSTRIES X SERVICES

- CONSTRUCTION & IMMOBILIER
- FINANCE & ASSURANCE
- CHIMIE
- ENERGIE & UTILITIES
- TRANSPORT & MOBILITÉS
- MÉDIAS & COMMUNICATION

- Performance énergie & carbone
- Résilience climatique
- Numérique & Com responsables
- IA de confiance**
- Finance responsable
- Prévention & Performance HSE

1 Centre d'innovation

CERTIFICATIONS & PARTENARIATS



Agenda

-
1. Contexte réglementaire et intérêts de la certification
 2. Retour d'expériences Axionable
 3. Conclusions & perspectives



Un consensus mondial s'est établi entre 2017 et 2020 autour des principes d'une IA digne de confiance, fondée sur 7 exigences clés ...

Principales sources



<>
Déclaration de Montréal
IA responsable_
</>

Consensus mondial autour des 7 piliers de l'IA de confiance

1. Contrôles humains (Human In The Loop)

L'IA sous le contrôle de l'humain et des processus comportant des points de contrôle et de débrayage humains.

2. Robustesse technique et sécurité

La fiabilité des algorithmes face à des cas aux limites, et les protection contre les risques de malveillance.

3. Respect de la vie privée et gouvernance des données

Le respect des réglementations sur les données à caractère personnel tels que le RGPD.

4. Transparence

La traçabilité, l'explicabilité et la communication de l'ensemble des prédictions selon les contraintes métiers et réglementaires.

5. Diversité, non-discrimination et équité

Détection des biais et ré équilibrage des données avec des processus statistiques permettant de générer des modèles plus équitables.

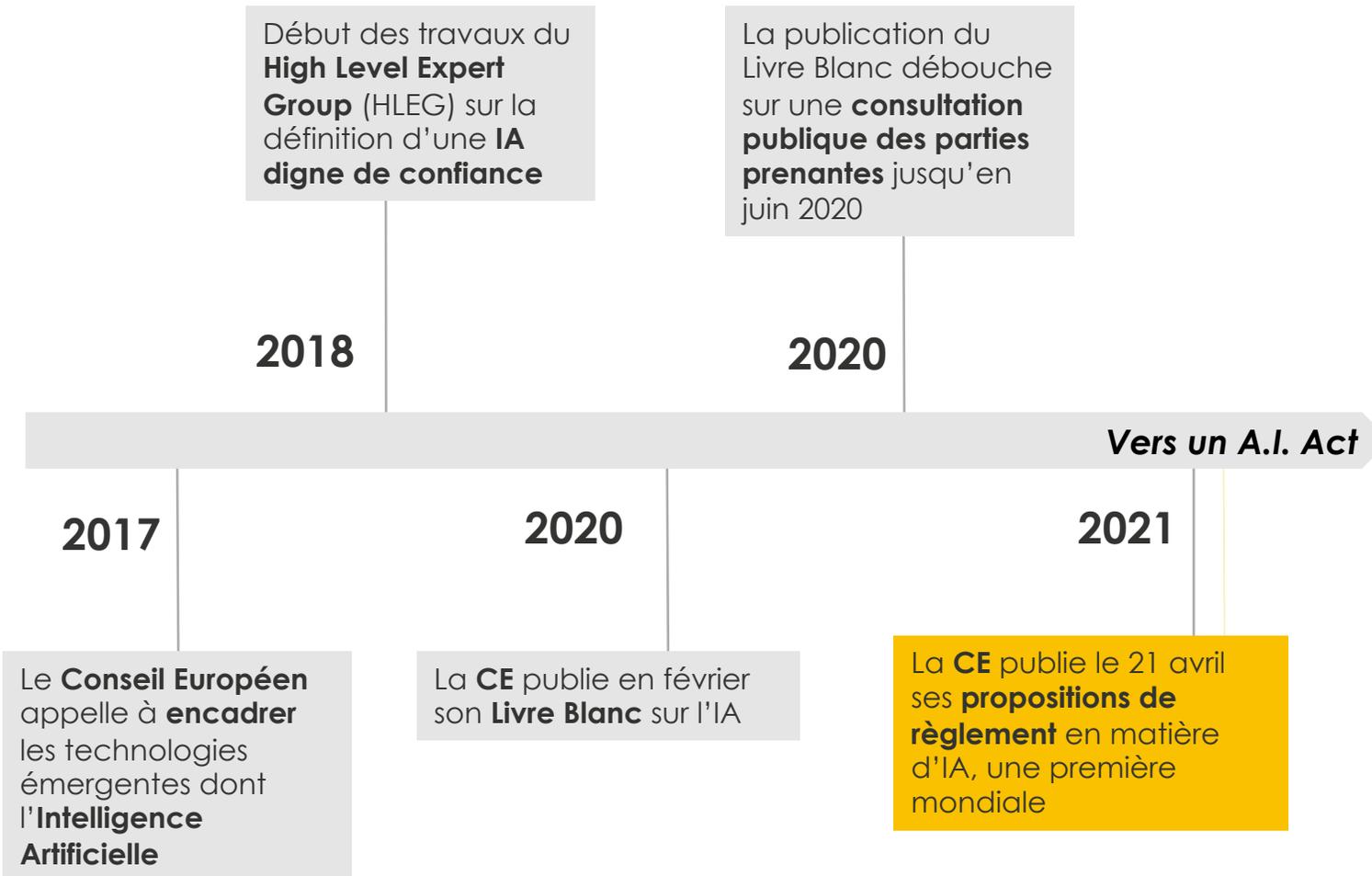
6. Bien-être environnemental et sociétal

Alignement d'intérêts business et environnementaux / sociétaux par l'optimisation des entraînements et des ressources.

7. Responsabilités associées au traitement

Une gouvernance adaptée au traitement réalisé par l'IA, intégrant toutes les parties prenantes.

... et le 21 avril 2021 marque un tournant avec le projet de réglementation européenne qui devrait entrer en application d'ici 18 à 24 mois sous la forme d'un AI Act...



Les objectifs du futur règlement européen de l'IA

 **21 avril 2021**



- 1 Garantir que les systèmes d'IA mis sur le marché de l'Union sont **sûrs** et respectent les **droits fondamentaux** et les **valeurs de l'Union**.
- 2 Dessiner un **cadre juridique harmonisé** pour faciliter l'investissement et l'innovation dans une IA digne de confiance.
- 3 Éviter la **fragmentation** du marché et des réglementations en matière d'IA.
- 4 Assurer la **position** de l'UE dans le **paysage mondial de l'IA** dominé jusqu'à présent par les mastodontes économiques que sont la Chine et les États-Unis.

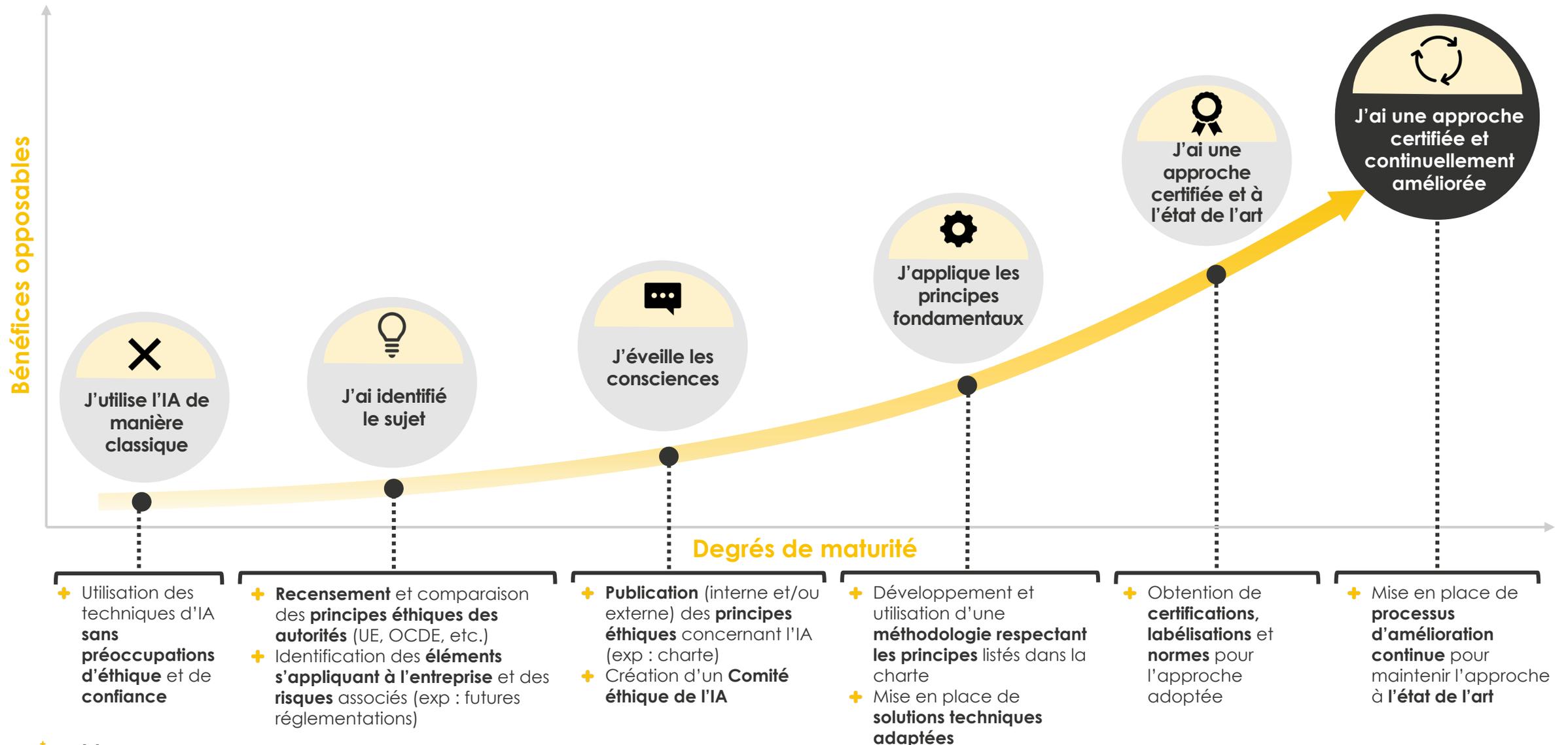
... et s'appliquerait de façon graduée selon 4 niveaux de risque posés sur la santé, la sécurité et les droits fondamentaux des utilisateurs...



Les 4 niveaux de risques de l'IA tels que définis par le projet de réglementation de la commission

Niveau de risque	4	3	2	1
Qualification	Risque inacceptable	Risque élevé	Risque limité	Risque Minimal
Qualification	Menace claire pour la sécurité, les moyens de subsistance et les droits des citoyens	Risques élevés pour la santé ou la sécurité des personnes	Interaction de personnes physiques avec une IA sans conséquences ni risque sur leur santé et leur sécurité	Risque minimal, voire nul, pour les droits ou la sécurité des citoyens
Règle	Interdiction	Obligations et process de validation strictes	Obligation de transparence	Pas de règle spécifique
Exemples	<ol style="list-style-type: none">1. Manipulation du comportement humain2. Notation des citoyens,3. identification biométrique en direct à distance à des fins répressives...	<ol style="list-style-type: none">1. Contribution à la sécurité de produits règlementés par l'UE2. Implication dans des systèmes déployés dans des secteurs d'activité à risque	<ol style="list-style-type: none">1. Chatbots2. Détection des émotions3. Catégorisation sociale4. Manipulation de contenus (deepfakes)	<ol style="list-style-type: none">1. Filtre chien sur Instagram2. Fonds d'écran zoom3. Etc.

Dans ce contexte, les entreprises se saisissent des enjeux de l'IA de confiance, avec différentes étapes de maturité et des degrés d'opposabilité des bénéfices



La certification IA LNE est le certificat disponible et mature aujourd'hui, il s'intègre dans un ensemble de labels/normes avec des maturités différentes

Panorama non exhaustif des initiatives de label / certificats / normes de l'IA (de confiance)



	Charte Arborus GEEIS-IA ⁽⁷⁾	GoodAlgo (ex Label ADEL d'ADELIAA) ⁽⁷⁾	Label IA Responsable et de Confiance ⁽¹⁾	Certification IA du LNE ⁽⁸⁾	Consultation nationale sur la stratégie de normalisation de l'IA ⁽⁹⁾	Focus Groupe IA du CEN / CENELEC ⁽⁹⁾	Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) ⁽³⁾	ISO/ IEC 42001 et TR 24028 ⁽⁶⁾
Objectif	<ul style="list-style-type: none"> S'assurer que les décisions automatisées fondées sur l'IA ne défavorisent pas les femmes, ou les minorités 	<ul style="list-style-type: none"> Proposer une labélisation des processus IA avec un audit systémique et une évaluation algorithmique des modèles d'IA 	<ul style="list-style-type: none"> Établir un référentiel d'évaluation des pratiques responsables dans le ML 	<ul style="list-style-type: none"> Proposer un référentiel « management process » co-construit avec les acteurs du marché et qui apporte des réponses à l'AI Act de l'UE 	<ul style="list-style-type: none"> Consultation des parties prenantes françaises pour définir la stratégie nationale de normalisation de l'IA 	<ul style="list-style-type: none"> Normaliser les définitions et les référentiels En cours de définition précise 	<ul style="list-style-type: none"> Proposer un programme de certification d'une IA transparente et responsable 	<ul style="list-style-type: none"> Standardiser les pratiques en matière d'IA afin de proposer des solutions d'IA de confiance
Périmètre	<ul style="list-style-type: none"> Éthique et Biais, avec un focus sur la fonction RH 	<ul style="list-style-type: none"> Robustesse, discrimination, fiabilité, qualité, interprétabilité, transparence et sécurité 	<ul style="list-style-type: none"> Privacy, Biais, Performance, Explicabilité, Responsabilité, Gestion du modèle, Suivi de l'impact SE 	<ul style="list-style-type: none"> Performance, gestion des risques, robustesse, explicabilité, éthique 	<ul style="list-style-type: none"> En cours de définition 	<ul style="list-style-type: none"> En cours de définition 	<ul style="list-style-type: none"> Auditabilité, Transparence, Biais 	<ul style="list-style-type: none"> Biais, Transparence, explicabilité, Privacy, Auditabilité
Disponibilité	<ul style="list-style-type: none"> Janvier 2021 	<ul style="list-style-type: none"> 2017 	<ul style="list-style-type: none"> Référentiel et labélisation depuis novembre 2021 	<ul style="list-style-type: none"> Référentiel depuis juillet 2021 Certification depuis septembre 2021 	<ul style="list-style-type: none"> Non communiqué 	<ul style="list-style-type: none"> Non communiqué 	<ul style="list-style-type: none"> Non communiqué 	<ul style="list-style-type: none"> Non communiqué

La combinaison des référentiels existants du LNE et de Labelia qui se complètent assure in fine un processus IA de confiance, bout en bout et « by design »

LNE et Labelia ce sont les initiatives que nous considérons les plus abouties, avec un référentiel et certificat/label disponible. De plus ces référentiels ont été co-construits par des groupes de travail (grands groupes , PME et data-scientists)



Et autres référentiels ou guide : ACPR, AI act, ImpactAI

1

Processus IA dont gestion des risques

Des **processus documentés, systématisés** sur l'ensemble des étapes du cycle de vie IA, avec des **contrôles** et outils de **gestions des risques**

Seuls, des processus avec des **angles morts sur la RSE**

+

2

IA responsable

Des **guides, bonnes pratiques et outils** pour une IA **respectueuse** des enjeux **humains** et **environnementaux**

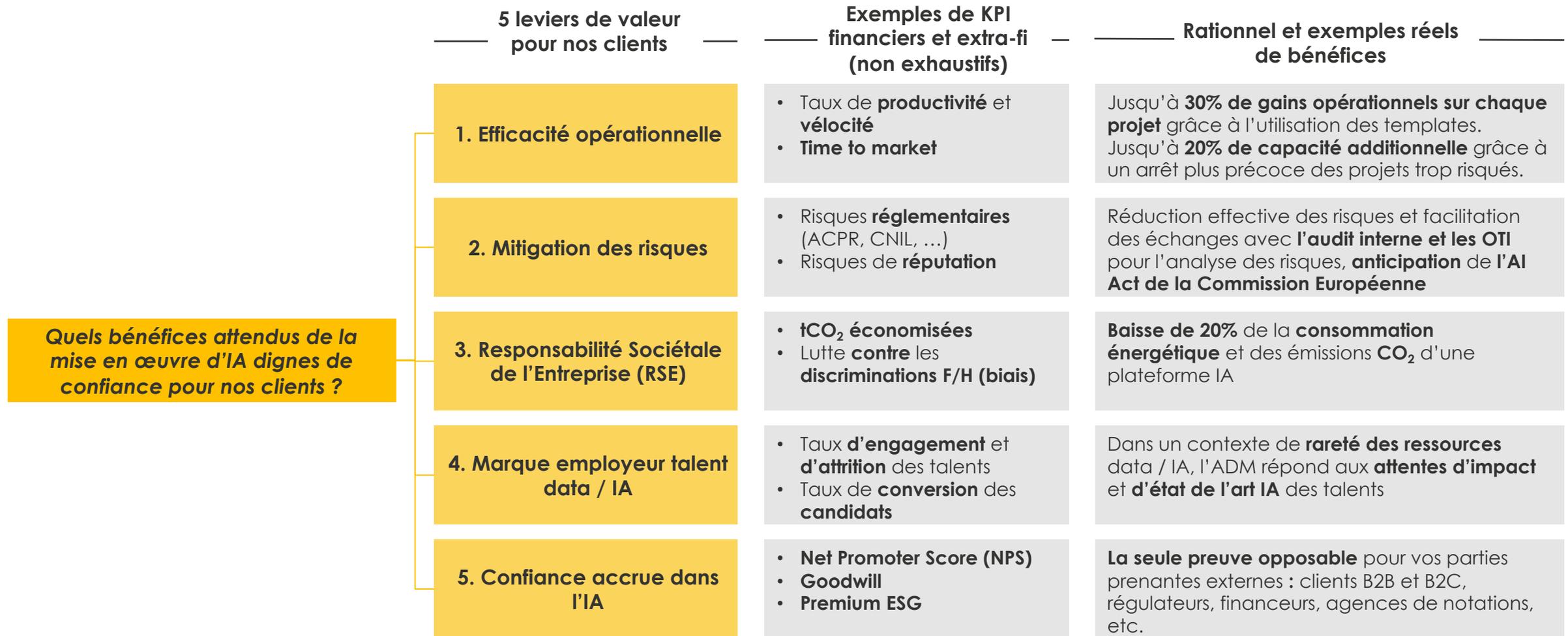
Seuls, des **intentions qui ne se concrétisent pas** et des risques qui se matérialisent

=

1. Processus IA dont gestion des risques + 2. IA responsable

Une IA digne de confiance, efficiente, dé-risquée et responsable

La mise en œuvre effective de ces recommandations permet à nos clients de viser des gains financiers et extra financiers matériels sur 5 leviers de valeur



Agenda

1. Contexte réglementaire et intérêts de la certification
2. Retour d'expériences Axionable
3. Conclusions & perspectives



La mise en conformité de notre méthodologie avec les référentiels nous a permis de faire évoluer notre méthodologie existante

Depuis début 2020

Initiation de notre méthodologie ADM pour l'IA de confiance, travaux R&D « Responsable ML »

T4 2020 – S1 2021

Contribution au GT de création du référentiel du LNE (12 ateliers, revue publique)

Sept-Nov 2021

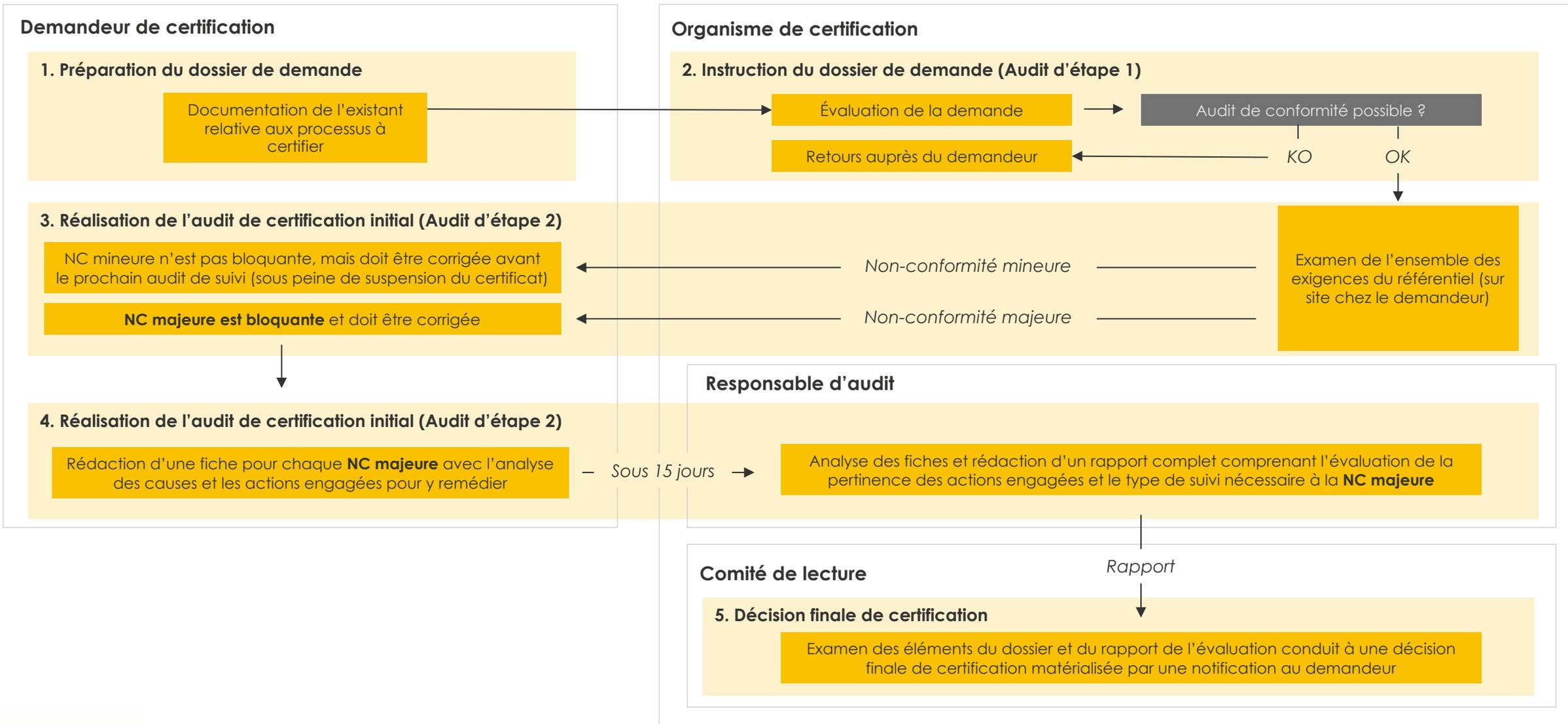
Mise en conformité avec le référentiel LNE et certification

Déc 2021 – Janv 2022

Mise en conformité avec le référentiel Labelia et labélisation

Nov 2022

Audit de suivi LNE (T + 1 an)



Sur un périmètre de 4 processus couvrant l'intégralité du **cycle de vie des IA**, le fournisseur d'IA est soumis à un ensemble d'**exigences** matérialisées par un travail de **documentation** et de communication auprès du client via une fiche produit.



1. Conception

Sur la phase de conception, le fournisseur d'IA doit notamment documenter les **attentes du client**, les **exigences réglementaires** et **normatives**, justifier l'**approche** retenue pour le développement de l'IA et réaliser une **analyse des risques** préliminaire.



2. Développement

Sur la phase de développement, le fournisseur d'IA doit notamment respecter des exigences en matière de **qualité des données**, décrire et justifier le **processus d'apprentissage** retenu et mettre à jour l'**analyse de risque** préliminaire.



3. Évaluation

Sur la phase d'évaluation, le fournisseur d'IA doit notamment mettre en œuvre un **protocole d'évaluation** strict en justifiant les métriques utilisées, identifier les **facteurs d'influence** sur la performance et les **biais**, évaluer la **robustesse** et la **résilience** du modèle et mettre à jour l'**analyse de risque** préliminaire.



4. Maintenance en condition opérationnelle

Sur la phase de MCO, le fournisseur d'IA doit notamment documenter toute **modification** relative à la fonctionnalité d'IA, inclure un **mécanisme de contrôle** de l'évolution de la performance pour l'utilisateur et organiser le **suivi du processus d'apprentissage** après déploiement de l'IA.

**REFERENTIEL DE CERTIFICATION
DE PROCESSUS POUR L'IA**
**Conception, développement, évaluation et
maintien en conditions opérationnelles**

Réf. rédacteur :
LNE/DEC/CITI/CH
LNE/DEC/IA/GA

Révision n°2.0

Approbation LNE : 12/07/2021



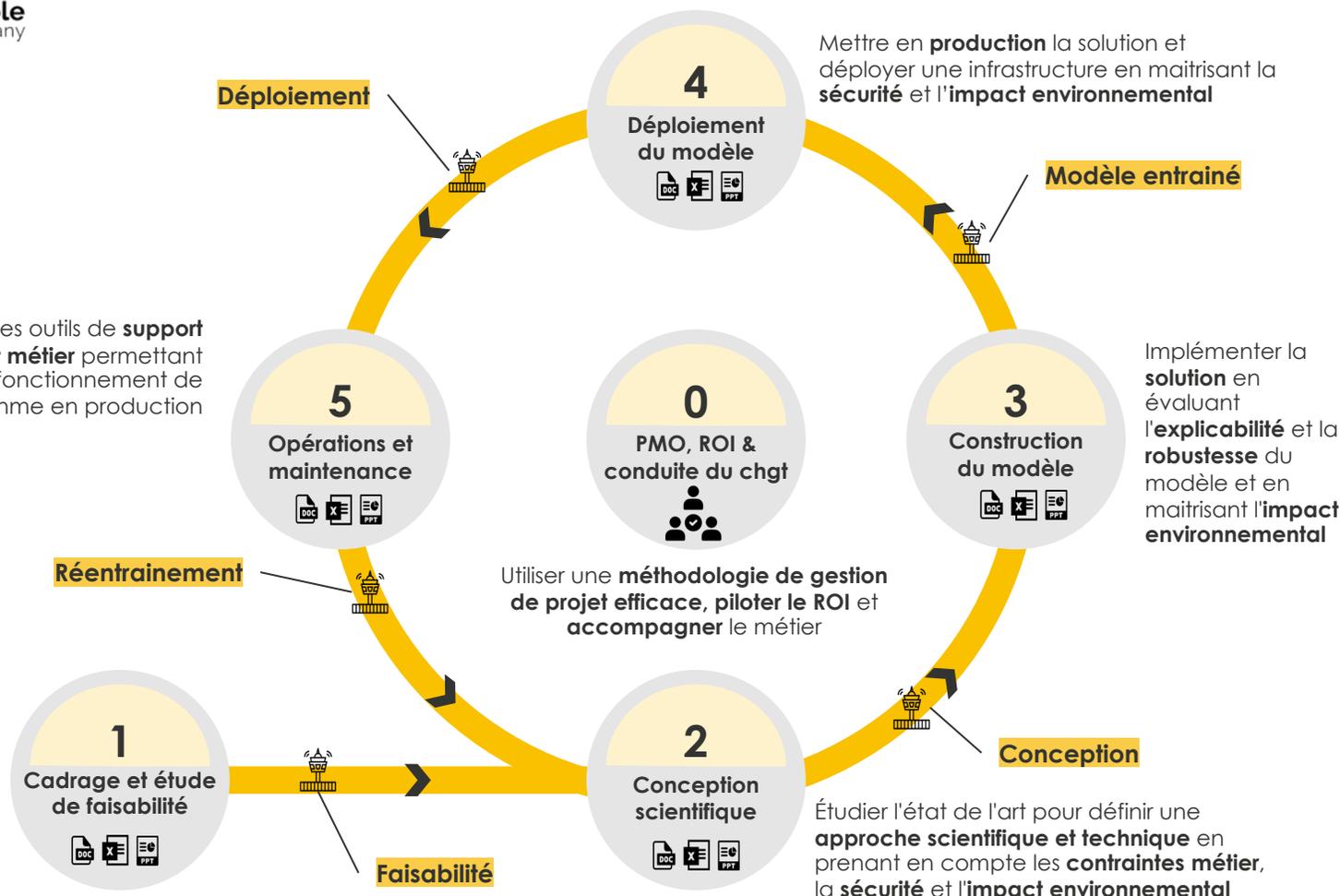
Processus certifié	Définition	Exemples de processus audités
Conception	Transformer une expression de besoin en spécifications fonctionnelles	Spécification et prise en compte des exigences normatives et réglementaires
Développement	Traduire ces spécifications en une version de la fonctionnalité d'IA prête à être évaluée	Apprentissage Qualité des Banques De Données.
Evaluation	Vérifier la conformité du système aux spécifications définies avant son déploiement	Définition des protocoles d'évaluation, des métriques, sur l'ensemble des outils d'évaluation qui permettent de rendre compte de l'efficacité de ces systèmes intelligents.
Maintien en conditions opérationnelles	Assurer la conformité de la fonctionnalité d'IA aux spécifications définies après son déploiement et tout au long de sa phase d'exploitation	Toutes les caractéristiques propres au maintien de ces conditions opérationnelles. Les systèmes d'IA peuvent évoluer tout au long de leur vie avec des dérives et des dégradations de performance.



Axionable Delivery Method (ADM) pour l'IA de confiance



Mettre en place des outils de **support technique et métier** permettant d'assurer le bon fonctionnement de l'algorithmes en production



Les outils de l'ADM

- Collaboration & workflow de validation
- Tour de contrôle pour GO / NOGO
- Templates de documentation

Projets audités :



Projet 1 : scoring ESG small / mid cap

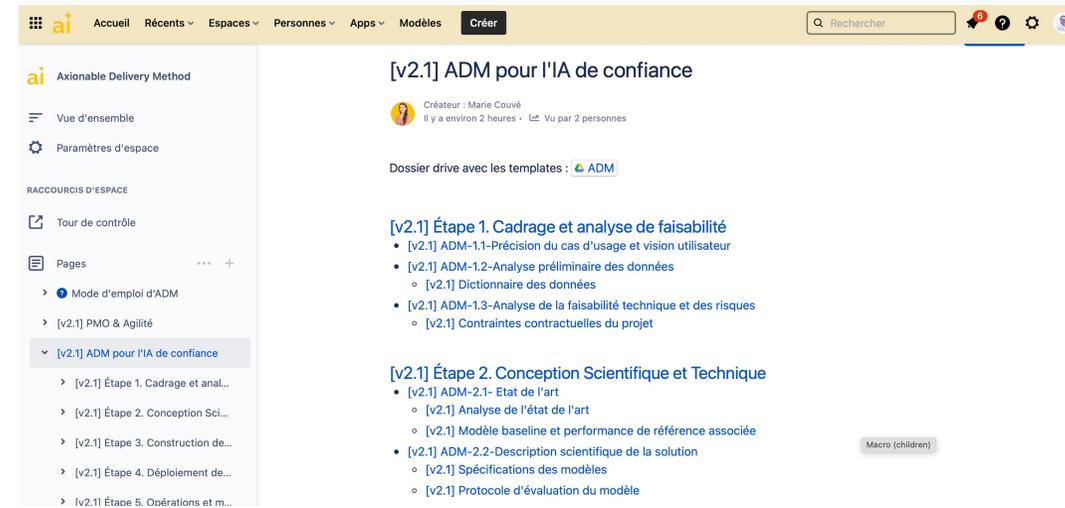


Projet 2 : prédiction climatique et sinistralité

Templates utilisés :

Template de vision utilisateur du produit IA

Template projet IA bout en bout sous Confluence



ai Accueil Récents Espaces Personnes Apps Modèles Créer

Rechercher

[v2.1] ADM pour l'IA de confiance

Créateur : Marie Couvé
Il y a environ 2 heures - Vu par 2 personnes

Dossier drive avec les templates : ADM

[v2.1] Étape 1. Cadrage et analyse de faisabilité

- [v2.1] ADM-1.1-Précision du cas d'usage et vision utilisateur
 - [v2.1] Dictionnaire des données
- [v2.1] ADM-1.2-Analyse préliminaire des données
 - [v2.1] Contraintes contractuelles du projet

[v2.1] Étape 2. Conception Scientifique et Technique

- [v2.1] ADM-2.1- Etat de l'art
 - [v2.1] Analyse de l'état de l'art
 - [v2.1] Modèle baseline et performance de référence associée
- [v2.1] ADM-2.2-Description scientifique de la solution
 - [v2.1] Spécifications des modèles
 - [v2.1] Protocole d'évaluation du modèle

Macro (children)

Points forts identifiés par les auditeurs

- **Très bonne préparation à l'audit** et une bonne implication des équipes pendant l'audit
- Les **rétrospectives de fin de projet** permettent l'amélioration continue
- Les **entretiens techniques standardisés**
- Le **template ADM sous Confluence, et les templates Word, PPT, Excel**

Des axes d'amélioration que nous avons pris en compte

- Mettre en place une **vision globale des tours de contrôle**
- **Automatiser systématiquement les séparations des jeux de données**
- **Formaliser les analyses d'impact et les communications au client**, notamment lors de modifications structurelles du projet
- **Justifier et documenter le choix des métriques** systématiquement
- Prévoir une **analyse de non régression systématique** lors de la mise à jour des modèles
- Corriger les Non-conformités mineures pour le prochain audit

Bilan de l'audit : pas de Non-conformités majeures, 5 non-conformités mineures
→ **Une certification assurée !**

En amont de l'audit :

- Choix des projets à auditer, et du périmètre (complet ou partiel)
- Mise en conformité de la documentation des projets avec le référentiel LNE
- Mise en conformité de la méthodologie ADM avec le référentiel LNE et définition des équivalents
- Envoi du dossier de demande de certification et échanges avec les auditeurs

≈ 80 jours de travail

Pendant l'audit :

- Présence des équipes pour répondre aux auditeurs pendant la semaine d'audit sur site : 4 personnes
- Mise à disposition de la documentation et du code pour l'audit
- Sollicitation d'autres interlocuteurs si nécessaire

Après l'audit :

- Réponses aux fiches de non-conformité
- Derniers échanges avec les auditeurs

≈ 20 jours de travail

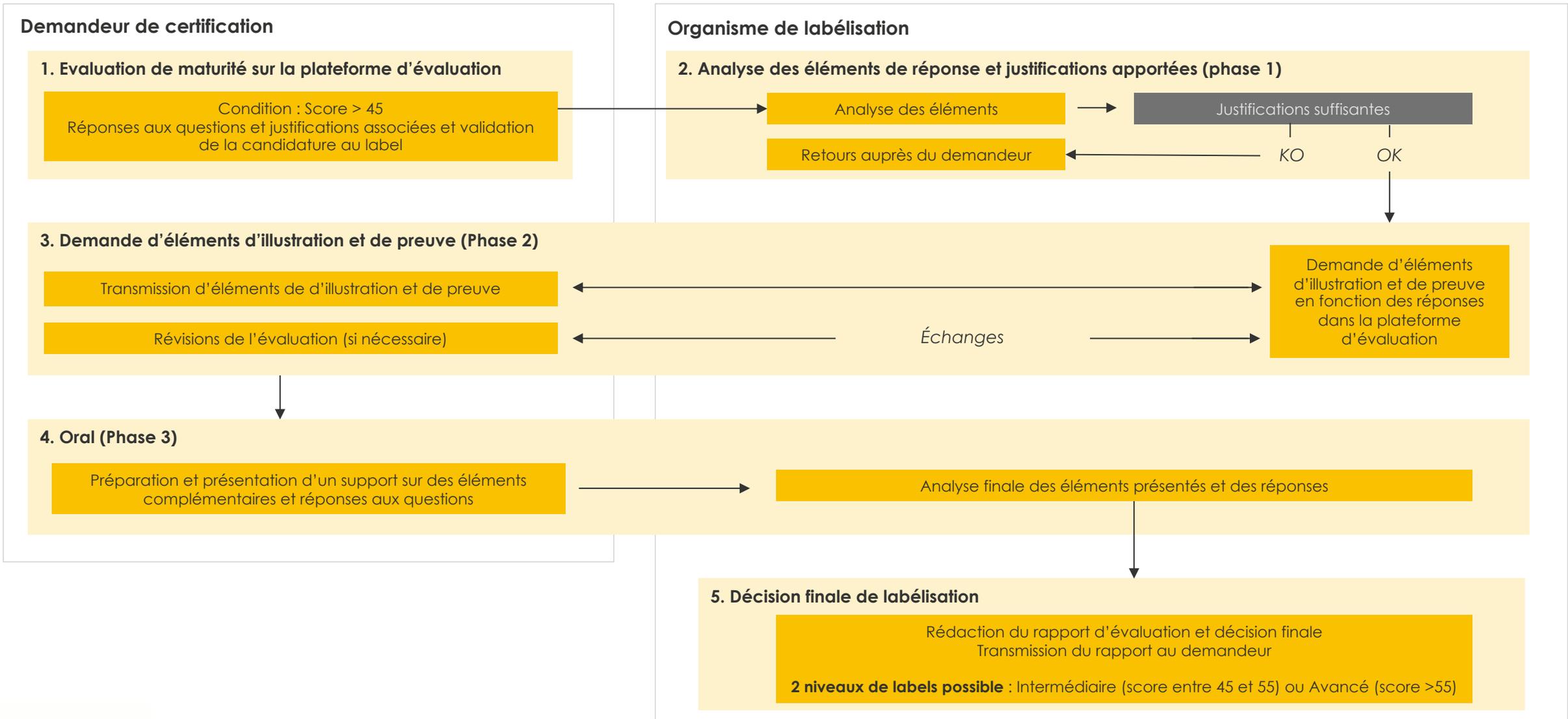
Coûts de la certification (pour Axionable)

- Audit initial : 10,5k€ HT
- Audit de suivi : 3,9k€ HT / an

Au total :

- ≈ 100 jours dépensés (pour l'audit)
- 5/6 personnes mobilisées
- 5 non-conformités mineures
- 2 projets audités
- **1 méthodologie certifiée**

Total ≈ 100 jours de travail
18k€ HT sur 3 ans



Points forts identifiés par les auditeurs

- **Haut niveau de maturité sur les enjeux et pratiques de l'IA responsable et de confiance**
- **Maturité « horizontale »** : connaissance et une maîtrise d'un très large ensemble de sujets qui compose le champ de l'IA responsable et de confiance dans la vision du référentiel cadre
- **Maturité « verticale »** : niveau d'expertise sur certains sujets pointus diverse selon les sujets
- **Politique d'exclusion commerciale** particulièrement engagée
- Grande diligence dans les réponses aux questions et la fourniture de documents de preuve ou d'illustration

Des axes d'amélioration que nous avons pris en compte

- Section 1 (Protection des données) avec un score moyen
- Systématiser la mise en place des **intervalles de confiance**
- Utiliser des techniques plus sophistiquées pour **l'anonymisation et minimisation des données**
- Systématiser les approches de « **fairness metrics** »
- D'autres éléments peuvent être améliorés pour augmenter le score obtenu (nous avons obtenu une note de 68,9 sur 100)

Bilan de l'audit : un score de 68,9
→ Labélisation niveau Avancé obtenue !!

Phase 1 – questionnaire :

- Prise de connaissance du référentiel et des ressources fournies
- Mise en conformité de notre méthodologie ADM avec le référentiel
- Evaluation sur la plateforme en ligne : réponse aux questions et justifications
- Révisions suite à l'analyse effectuée par Labelia
- Échanges avec l'équipe Labelia

≈ 15 jours de travail

Phase 2 - éléments d'illustration et de preuve :

- Envoi des documents demandés par Labelia avec des explications
- Révisions suite à l'analyse effectuée par Labelia

Phase 3 – oral :

- Présentations des éléments suggérés par Labelia (différents des documents de la phase 2)
- Échanges et précisions par rapport aux éléments apportés

≈ 3 jours de travail

Coûts de la labélisation (pour Axionable)

- Adhésion annuelle : 3000€ HT

Au total :

- ≈ 18 jours dépensés
- 2 personnes mobilisées
- Score : 68,9
- 1 entreprise labélisée niveau avancé

Total ≈ 18 jours de travail
3000€ HT par an

Agenda

1. Contexte réglementaire et intérêts de la certification
2. Retour d'expériences Axionable
3. Conclusions & perspectives



Pour rester à l'état de l'art et conserver nos labels et certificats, nous nous adaptons à l'évolution des référentiels et initiatives

Veille réglementaire et technologique

- Suivi des évolutions réglementaires comme l'AI Act
- Suivi des réglementations IA sectorielles : ACPR, ASN, etc.
- Identifications des nouvelles initiatives d'IA de confiance
- Suivi et test des outils (open-source, software...) à l'état de l'art

Amélioration continue de notre méthode et nos outils

- Adaptation aux nouveaux référentiels jugés pertinents
- Mise à jour en fonction des évolutions des référentiels
- Audit de suivi et renouvellement de nos certificats et labels
- Intégration à notre méthodologie des nouveaux outils pertinents
- Adaptation de notre méthodologie aux référentiels sectoriels (ACPR , ASN...) et propres à l'entreprise (RGPD, cyber sécurité...)

Formation de nos équipes

- Formation continue des équipes à notre méthodologie ADM pour l'IA de confiance
- Formation aux outils d'IA de confiance
- Partage et retours d'expériences en interne

SAVE THE DATE

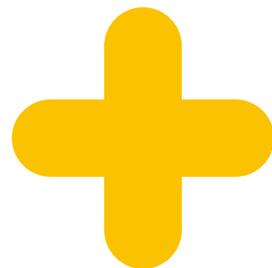
datacraft* &  **IMPACT AI**
organisent

AI ACT DAY

KEYNOTES - TABLES RONDES - WORKSHOPS

 **JEUDI 15 décembre 2022**
09h00 - 19h00

 **datacraft*** Sorbonne Center for Artificial Intelligence
4 Place Jussieu, 75005 Paris



Vos contacts

Gwendal Bihan
Président & CTO

gwendal@axionable.com
+33 7 84 10 16 50

José Sanchez
**Senior Manager ML
Engineer**

jose@axionable.com
+33 6 16 44 25 47