AEXIA™ FRAMEWORK V3.1

Guide d'utilisation

Mettre en œuvre AEXIA™ pas à pas : inventaire, niveaux, portes, mesures, pilotage

---

Mohamadou SIDIBE
AI BIAIS INITIATIVE
Chief Data Officer, Expert en gouvernance des données et de l'IA
Aligné sur AEXIA™ Framework V3.1 · Licence CC BY 4.0

Comment lire ce guide

Ce guide explique **comment utiliser AEXIA™ Framework V3.1 au quotidien** : qui fait quoi, dans quel ordre, avec quels documents et quelles preuves. Il s'adresse aux CDO, DPO, Chief AI Officers, data stewards, responsables risques et auditeurs. Il ne remplace pas le Framework : chaque définition et chaque règle renvoient à sa section.

Le kit documentaire

Document	À quoi il sert	Quand l'utiliser
AEXIA™ Framework V3.1	Référence : concepts, règles, mesures	En cas de doute sur une règle
Guide d'utilisation (ce document)	Mode d'emploi pas à pas	Au démarrage et à chaque cycle
Grille d'évaluation AEXIA-O	Mesurer la maturité de l'organisation, avec preuves	1 à 2 fois par an
Fiche d'auto-évaluation	Synthèse du triptyque, alertes, indicateurs, RAD-lite	Chaque trimestre ou semestre
AEXIA Card (gabarit)	Fiche d'identité d'un système d'IA	Un exemplaire par système, mis à jour à chaque porte et incident
AEXIA Card (exemple Meridian)	Exemple rempli, système fictif	Pour former les équipes

> **Statut à garder en tête**
> AEXIA™ est une proposition de l'auteur, publiée en accès ouvert. Les mesures DCME, ECS et RAD ont fait l'objet d'une validation **préliminaire sur données synthétiques et petits jeux publics** : elles n'ont pas encore été validées sur systèmes réels ni relues par des pairs. Les seuils cités dans les documents sont des **repères de politique**, pas des normes. AEXIA n'est ni une certification ni un label pour l'instant.

1. AEXIA™ en dix minutes

AEXIA™ répond à une question que les référentiels existants (NIST AI RMF, ISO/IEC 42001, MLTRL, AI Act) ne traitent pas frontalement : **qui est responsable d'une décision produite par un système apprenant, et sur quelles preuves ?** Il les orchestre au lieu de les remplacer.

Deux axes, un croisement

- **Axe O (organisation)** : six piliers (Fondements, Architecture, Gouvernance, Conformité, Opérations, Mesure et amélioration) évalués de N0 à N5. Le niveau publié est celui du pilier le plus faible.
- **Axe S (système)** : chaque système reçoit un niveau de S0 à S9 (dix niveaux), aligné sur les MLTRL. Il est réévalué à chaque porte et à chaque incident.
- **Le croisement O × S** dit ce que l'organisation a le droit de déployer, et avec quels contrôles.

Les quatre mécanismes

Mécanisme	Ce qu'il impose
Maillon faible	Le niveau publié est celui du composant le plus faible, jamais une moyenne.
Portes probantes	Aucun passage de niveau par déclaration : un jeu fermé d'artefacts, examiné en revue.
Rétrogradation	Un système qualifié peut redescendre : dérive, incident, changement de population, évolution réglementaire, réentraînement.
Proportionnalité	L'effort est proportionné à la classification de risque AI Act.

Ce que l'on publie : le triptyque

Valeur	Définition	Usage
Niveau AEXIA-O	Niveau du pilier le plus faible	Publié en externe
Profil AEXIA-O	Niveaux des six piliers (radar)	Pilotage interne : où investir
Distribution AEXIA-S	Répartition du parc par niveau S0 à S9	Arbitrages opérationnels

La matrice de décision

Org.	Systèmes S0-S5	Systèmes S6-S7	Systèmes S8-S9
N0-N1	Toléré : phase exploratoire	**Interdit** : pas de contrôles	**Alerte majeure** : production sans gouvernance
N2	Toléré avec traçabilité	Autorisé sous condition : porte franchie en comité	Autorisé pour risque limité ; haut risque : non
N3	Normal	Normal, porte franchie sur preuve	Autorisé, y compris haut risque, sous surveillance mesurée
N4-N5	Normal	Normal, automatisé	Normal, rétrogradation automatisée sur dérive

2. Le parcours en six étapes

Le cycle de vie AEXIA™ (MAP, DEFINE, BUILD, GOVERN, OPERATE, IMPROVE) se traduit en six étapes de travail. Les durées sont celles indiquées par le Framework pour une organisation qui démarre, elles varient avec la taille du parc.

Phase	Ce que vous faites	Livrables	Durée indicative
MAP	Inventorier les systèmes, les classer (AI Act), positionner chacun sur S0-S9	Inventaire, classification, niveau initial, premières AEXIA Card	2 à 4 semaines
DEFINE	Fixer politique, rôles (RACI), méthodes, seuils de politique	Charte, RACI, matrice de sélection SHAP/LIME/DCME	4 à 8 semaines
BUILD	Déployer outils, registres, tableau de bord, former	Pipeline MLOps augmenté, catalogue, tableau de bord	2 à 6 mois
GOVERN	Instaurer les portes, le DGC, la piste d'audit	Décisions de porte, rapport DGC, rapport de conformité	En continu
OPERATE	Gérer explications, incidents, niveaux de service, RAD	Procédures, niveaux de service, tickets	En continu
IMPROVE	Réévaluer la maturité, traiter les écarts	Score mis à jour, feuille de route, rapport au comité exécutif	Trimestriel

Les 90 premiers jours

1. Désigner les rôles minimaux (Data Owner, Data Steward XAI, DPO, sponsor CDO). 2. Inventorier tous les systèmes d'IA, même ceux « de test ». 3. Classer chaque système au sens de l'AI Act et noter la justification. 4. Remplir une première AEXIA Card pour les systèmes à haut risque. 5. Évaluer l'organisation avec la grille AEXIA-O. 6. Présenter le triptyque au comité exécutif.

3. Étape 1 : Inventorier et classer

3.1 Inventorier

Un système d'IA entre dans l'inventaire dès qu'il produit ou influence une décision qui concerne des personnes : score, tri, recommandation, détection, génération. Inclure les systèmes achetés, ceux des fournisseurs et les prototypes accessibles à des utilisateurs réels. Pour chacun, noter un code interne, le propriétaire métier, la finalité, les personnes concernées et la date de première mise en service.

3.2 Classer au sens de l'AI Act

1. Le système relève-t-il d'une pratique interdite (art. 5) ? Si oui : **inacceptable**, arrêt.
2. Est-il un composant de sécurité d'un produit réglementé (annexe I) ou relève-t-il d'un domaine de l'annexe III (crédit, emploi, éducation, accès aux services essentiels, justice, etc.) ? Si oui : **haut risque**.
3. Interagit-il avec des personnes ou génère-t-il du contenu soumis à transparence (art. 50) ? Si oui : **risque limité**.
4. Sinon : **risque minimal**.

Consigner la justification et la date : la fiche en a besoin, et le régulateur la demandera. Selon la section 9 du Framework (état du droit au 22 octobre 2026), les obligations des systèmes à haut risque de l'annexe III s'appliquent à compter du 2 décembre 2027 et celles des systèmes intégrés à des produits à compter du 2 août 2028 ; cette section se revoit chaque semestre.

3.3 Proportionner l'effort

Le Framework pose le principe de proportionnalité (mécanisme 4) sans détailler le contenu. Le tableau suivant est une **recommandation de ce guide**, à adapter par politique.

Classe	Effort	Contenu minimal
Minimal	Allégé	Inventaire ; sections 1 et 2 de la fiche ; revue annuelle
Limité (art. 50)	Standard	Fiche sections 1 à 4 et 6 ; méthode d'explication choisie ; revue semestrielle
Haut risque / interdit	Complet	Fiche complète ; portes S5→S6, S7→S8, S8→S9 ; DCME si variables protégées ou proxys ; ECS et RAD ; surveillance (art. 72)

4. Étape 2 : Positionner chaque système sur l'axe S

Le niveau S d'un système n'est pas un jugement : c'est **le plus haut niveau dont les preuves existent**, celui de son composant le moins mûr. En cas de doute, retenir le niveau inférieur.

S	Étape	Preuves typiques	Environnement
S0	Premiers principes	Idée, revue de littérature, bases mathématiques	Aucun
S1	Recherche orientée	Expérimentations exploratoires consignées	Laboratoire
S2	Preuve de principe	Concept fonctionnel sur cas restreint	Laboratoire
S3	Développement système	Code structuré, pipeline reproductible	Laboratoire
S4	Preuve de concept	Validation sur données réelles	Représentatif
S5	Capacité démontrée	Capacité métier identifiable et mesurée	Représentatif · porte critique
S6	Développement applicatif	Qualité logicielle de production ; registre des explications	Représentatif
S7	Intégration	Intégration aux systèmes de production	Opérationnel restreint
S8	Qualifié	Dossier de conformité complet ; décision formelle du DGC	Opérationnel
S9	En service surveillé	Exploitation, maintenance, surveillance active (art. 72)	Opérationnel

4.1 Les quatre portes

Porte	Artefacts exigés	Décisionnaires	Où la tracer
S2 → S3	Fiche de risque AI Act, classification, revue de la préparation des données	Data Owner + Data Steward	Fiche, section 3
S5 → S6	Justification de la méthode d'explicabilité, seuils de qualité, quantification des risques	DGC (porte critique)	Fiche, sections 3 et 5
S7 → S8	Dossier de conformité complet, jeu de référence figé, procédure de surveillance	DGC + DPO	Fiche, sections 3, 6 et 8
S8 → S9	Pipelines de surveillance actifs, seuils de dérive, procédure de rétrogradation testée	DGC + DSI	Fiche, sections 3 et 7

Conduite d'une porte. (1) Le Data Steward assemble les artefacts et référence le dossier dans la fiche. (2) Les décisionnaires examinent en revue, jamais par signature isolée. (3) La décision (franchie ou refusée) est consignée avec la date, l'instance et le commentaire. (4) Une porte refusée reste refusée : un plan de remédiation et une nouvelle revue sont nécessaires. (5) Le niveau n'est mis à jour qu'après la décision.

4.2 Fixer le niveau attendu S_att

Le RAD compare le niveau attendu et le niveau effectif. Le Framework ne fixe pas S_att : il doit être **déclaré par politique et publié**. Repère suggéré par ce guide : S9 pour un système à haut risque en exploitation, S8 avant sa mise en production et le niveau correspondant au stade prévu pour un système en développement.

5. Étape 3 : Évaluer l'organisation sur l'axe O

L'évaluation AEXIA-O se fait avec la grille (6 piliers, critères numérotés, preuve par critère). Elle prend une à deux journées pour un premier passage.

1. **Préparer les preuves** : chartes, RACI, procès-verbaux du DGC, catalogues, rapports d'audit, tableaux de bord.
2. **Évaluer pilier par pilier** : cocher un critère seulement s'il est prouvé ; un critère partiel compte comme non satisfait.
3. **Attribuer le niveau** : le plus haut niveau dont tous les critères, et ceux des niveaux inférieurs, sont satisfaits.
4. **Faire relire par un second évaluateur** si le résultat est communiqué à l'extérieur ; noter les désaccords.
5. **Publier le niveau du pilier le plus faible** et utiliser le profil pour décider des investissements.

> **Pourquoi pas un score moyen ?**
> Avec les pondérations F 20 %, A 15 %, G 20 %, C 20 %, O 15 %, M 10 %, une organisation dont la Conformité est à N0 et dont les cinq autres piliers sont à N4 obtient 3,2, soit « Géré », alors qu'elle est en exposition juridique directe. Le maillon faible affiche N0 et rend l'exposition visible. La grille propose le score pondéré comme repère interne facultatif, jamais publié seul.

Preuves usuelles par pilier

Pilier	Exemples de preuves acceptables
F · Fondements	Glossaire validé, benchmark de méthodes, matrice de sélection, rapport DCME avec graphe publié, standards de qualité des explications
A · Architecture	Schéma du pipeline MLOps, catalogue de données, registre des explications, documentation de l'API, configuration de la porte automatique
G · Gouvernance	Charte signée, RACI, procès-verbaux trimestriels du DGC, décisions de porte, rapport RAD-lite
C · Conformité	Registre des risques IA, classification de chaque système, correspondance RGPD / AI Act, rapport d'audit annuel
O · Opérations	Runbook XAI, accords de niveau de service, tickets d'incident, indicateurs opérationnels
M · Mesure	Score annuel, tableau de bord trimestriel, rapports ECS conformes aux règles de publication, plans d'amélioration

6. Étape 4 : Remplir l'AEXIA Card

Une fiche décrit un système. C'est le document qu'un régulateur demandera en premier : elle doit se suffire à elle-même et renvoyer aux dossiers de preuves. L'objectif de l'étude préenregistrée est de remplir une première fiche en **quatre heures ou moins** (médiane), c'est un objectif à vérifier, pas un résultat acquis.

Section de la fiche	Que renseigner	Piège à éviter
En-tête	Identifiant AEXIA-XXXXXXXX, version x.y.z, statut	Réutiliser un identifiant ; ne pas versionner
1 Identification	Nom, code, finalité, personnes concernées, rôles (fonctions plutôt que noms)	Finalité floue ; oublier la décision supportée
2 Classification	Catégorie, annexe, justification, date d'application	Classer sans justifier
3 Niveau S	Niveau actuel, S_att, historique des portes avec dossiers	Niveau déclaré sans dossier
4 Maillon faible	Composant déterminant, plan, échéance ; niveau O et case de la matrice	Plan sans responsable ni date
5 Méthode et mesures	Méthode ; blocs DCME, ECS et RAD avec leurs règles de publication	ECS sans ε, sans témoins, sans fidélité
6 Données	Niveau de préparation, jeu figé, lignée, biais connus (référence TriAxis)	Jeu de référence non figé
7 Surveillance	Seuils pour chacun des cinq déclencheurs, fréquence, dernier contrôle	Surveiller sans seuil de rétrogradation
8 Conformité	Pièces, audit, écarts, échéance, information des personnes	Écarts ouverts non datés
9 à 11	Historique de la fiche, validations, notes	Signatures sans rôle

6.1 Versionner la fiche

- **Majeur (x)** : changement de niveau S, de classification ou de méthode d'explicabilité.
- **Mineur (y)** : mise à jour des mesures (ECS, RAD), des données ou de la surveillance.
- **Correctif (z)** : correction de forme sans effet sur le contenu.

6.2 Valider le format

Le Framework prévoit un schéma JSON (aexia_card_schema.json, Draft-07) et un validateur Python. Les champs obligatoires sont : id, version, system_name, level_s, xai_method, data_readiness, compliance.

```
{ "id": "AEXIA-MERSCR01", "version": "0.3.0",
  "system_name": "Meridian consumer-credit scoring",
  "classification_ai_act": "haut_risque", "level_s": 5,
  "xai_method": "SHAP",
  "data_readiness": {"level": 4}, "surveillance": {},
  "compliance": {"status": "partial"} }
```

Limite du schéma actuel : il n'admet qu'une seule valeur pour xai_method, alors qu'un système peut combiner SHAP et DCME, et il ne contient pas les champs ECS et RAD de la fiche V3.1.

7. Étape 5 : Appliquer la matrice et décider

1. Prendre le **niveau AEXIA-O publié** de l'organisation.
2. Prendre le **niveau S** du système (fiche, section 3).
3. Lire la case de la matrice et la consigner dans la section 4 de la fiche.
4. Si la case est **Interdit** ou **Alerte majeure**, escalader au DGC et au CDO sans attendre le prochain comité.
5. Si la case est **Autorisé sous condition**, vérifier que la porte a été franchie en comité et que la condition est écrite.

Trois lectures types

Situation	Case	Décision attendue
Organisation N2, système à haut risque en S7	Autorisé sous condition	Porte franchie en comité, condition écrite (périmètre restreint, plan vers S8)
Organisation N1, système en S6	Interdit	Ne pas déployer : monter l'organisation à N2 ou ramener le système en laboratoire
Organisation N3, système à haut risque en S9	Autorisé sous surveillance mesurée	Maintenir surveillance, seuils et procédure de rétrogradation à jour

8. Étape 6 : Surveiller, rétrograder, revenir

Un système d'IA ne reste pas qualifié : les données dérivent, les populations changent, les modèles sont réentraînés. AEXIA prévoit cinq déclencheurs de rétrogradation.

Déclencheur	Effet	Obligation pour revenir
Dérive de données ou de concept au-delà du seuil	S9 → S6	Réexamen de la méthode, recalibrage, nouvelle porte S7 → S8
Incident d'explicabilité	S9 → S7	Analyse de cause racine, correction, revue du DGC
Changement significatif de population	S9 → S6	Revue de la préparation des données, contrôle d'équité, réexamen de la conformité
Évolution réglementaire applicable	S9 → S8	Mise à jour du dossier de conformité, nouvelle décision formelle
Modification du modèle (réentraînement)	S9 → S7	Comparaison au jeu de référence, validation des explications

Procédure en six temps

6. **Détecter** : un indicateur franchit son seuil de politique, ou un incident est déclaré.
7. **Qualifier** : identifier lequel des cinq déclencheurs s'applique.
8. **Rétrograder** : mettre à jour le niveau S dans la fiche (version majeure) et relire la case de la matrice.
9. **Décider** : si la nouvelle case est Interdit ou si le risque l'exige, suspendre ou restreindre l'usage.
10. **Corriger** : mener l'obligation de retour correspondant au déclencheur.
11. **Re-franchir la porte** : nouvelle décision en revue, consignée dans la fiche.

Les seuils de détection (par exemple un indice de stabilité de population, une baisse de performance, une dégradation de l'ECS sur des données récentes) sont **fixés par politique** et consignés dans la section 7 de la fiche. La surveillance après mise sur le marché (art. 72) doit faire l'objet d'un plan pour les systèmes à haut risque.

9. Utiliser DCME, ECS et RAD sans se tromper

Les trois mesures sont l'apport original de l'auteur Mohamadou SIDIBE, elles sont **préliminaires**. Leur emploi suit des règles de publication strictes : sans elles, un chiffre peut tromper. Les exemples numériques sont ceux de l'étude de cas Meridian (données synthétiques).

9.1 DCME : par où l'influence passe

- **Quand** : système à haut risque pour lequel une caractéristique protégée (ou un proxy) peut agir sur des variables que le modèle utilise. Sinon, SHAP suffit.
- **Procédure** : construire et publier le graphe causal avec sa justification ; estimer ϕD (direct) et ϕC (causal) ; calculer ψ = ϕC − ϕD ; vérifier l'efficacité et Σψ = 0 (tolérance 10^{-4}) ; mener l'analyse de sensibilité au graphe.
- **Lecture** : ϕD répond à « le modèle utilise-t-il la variable ? », ϕC à « la variable influence-t-elle le score, par tous les chemins ? ». ψ montre où l'influence est redistribuée. Sur Meridian : contribution directe du groupe nulle, contribution causale de +0,95 point.
- **Si l'identifiabilité n'est pas défendable** (confusion non observée possible) : publier ϕD seule et déclarer la limite ; ne pas parler de causalité.

9.2 ECS : l'explication est-elle stable ?

- **Quand** : pour toute explication destinée à un client, un régulateur ou un audit.
- **Règles de publication** : rapporter ε, N, la normalisation et une mesure de fidélité ; inclure un témoin aléatoire (plancher proche de 0) et un témoin constant (plafond trompeur de 1) ; donner plusieurs rayons (0,05 ; 0,1 ; 0,25 σ) ; publier l'intervalle de confiance bootstrap ; fixer par politique un seuil minimal.
- **Lecture** : l'ECS mesure la stabilité, jamais la vérité ni la qualité. Sur Meridian : 0,90 à ε = 0,1 σ, fidélité 0,71, témoins 0,00 et 1,00. Comparer aussi l'ECS par sous-groupe.

9.3 RAD : où et depuis quand est-on en retard ?

- **Quand** : pour piloter un portefeuille de systèmes devant le comité exécutif.
- **Calcul** : RAD-lite = Σ β × max(0, S_att − S_eff) / Σ β × S_att, entre 0 et 1. Rapporter **à part** l'ancienneté de l'écart le plus ancien et la durée moyenne d'ouverture.
- **Lecture** : une part pondérée du chemin restant à parcourir. Regarder la contribution de chaque système : sur un portefeuille de quatre systèmes (exemple Meridian), RAD-lite = 36 / 81 = 0,444, dont 89 % portés par deux systèmes.
- **Précautions** : aucun seuil normatif avant calibrage ; niveaux S ordinaux (lire l'écart en portes manquantes) ; une variation de quelques centièmes peut être du bruit de mesure ; variante cumulative facultative, avec T½ déclaré.

10. Piloter et rendre compte

Le pilotage tient sur une page : le triptyque, le RAD-lite avec l'ancienneté, les alertes de la matrice, les portes de la période et les décisions demandées au comité.

Objet	Cadence	Instance
Niveau S d'un système	À chaque porte et à chaque incident	DGC (selon la porte)
Évaluation AEXIA-O	Une à deux fois par an	CDO, DGC
Indicateurs, RAD-lite, alertes	Trimestrielle (tableau de bord)	DGC puis comité exécutif
Section réglementaire	Semestrielle	DPO, CDO
Feuille de route d'amélioration	Trimestrielle (phase IMPROVE)	CDO

Contenu type du rapport au comité

- Niveau AEXIA-O publié, profil des six piliers et évolution depuis la dernière évaluation.
- Distribution AEXIA-S, dont systèmes à haut risque, et alertes de la matrice.
- RAD-lite, contribution de chaque système, ancienneté des écarts.
- Portes franchies, refusées, rétrogradations et incidents de la période.
- Décisions demandées : arbitrages d'investissement, autorisations sous condition, échéances AI Act.

11. Rôles et responsabilités

Répartition proposée, adaptée de la section 4.3 du Framework (A = redevable, R = réalise, C = consulté, I = informé). À valider dans la charte de l'organisation.

Activité	CDO	DSI	DPO	Data Owner	Data Steward	DGC
Inventaire des systèmes	A	C	C	R	R	I
Classification AI Act	A	C	R	R	C	I
Niveau S et remplissage de la fiche	I	C	C	A	R	I
Porte S2 → S3	I	I	C	A/R	R	I
Porte S5 → S6	I	C	C	C	R	A
Porte S7 → S8	I	C	R	C	R	A
Porte S8 → S9	I	R	C	C	R	A
Mesures DCME, ECS, RAD	I	C	C	A	R	I
Surveillance et rétrogradation	I	R	C	A	R	I
Évaluation AEXIA-O	A	C	C	C	R	C
Rapport au comité exécutif	A/R	C	C	I	C	C

12. Cas pratique : la banque Meridian

Meridian est une banque fictive. Son score de crédit n'utilise pas la variable sensible, mais les dossiers du groupe protégé sont acceptés moins souvent (rapport des taux d'environ 0,83). Voici le parcours AEXIA complet, sur données synthétiques, la fiche remplie correspondante est fournie dans le kit.

Étape	Résultat
1. Inventaire	Score de crédit à la consommation MER-SCR-01, en mode ombre depuis le 1er septembre 2026
2. Classification	Haut risque, annexe III (solvabilité des personnes physiques) ; échéance 2 décembre 2027 ; effort complet
3. Niveau S	S5 ; S_att = S8 ; porte S5 → S6 refusée (seuils de qualité et quantification des risques non finalisés)
4. Matrice	Organisation N2, système S5 : toléré avec traçabilité (pas de décision appliquée aux clients)
5. DCME	Contribution directe du groupe : 0. Contribution causale : +0,95 point. Sans les arêtes issues du groupe, 28 % de l'effet est retrouvé : le graphe doit être justifié
6. ECS	0,90 à ε = 0,1 σ ; témoins 0,00 et 1,00 ; fidélité 0,71 ; écart de 0,03 entre groupes (moins stable pour le groupe protégé)
7. RAD	Système : 12 sur 32 ; portefeuille de quatre systèmes : 36 / 81 = 0,444 ; ancienneté 24 mois
8. Décision	Autoriser seulement sous conditions : décision documentée sur la légitimité des chemins d'influence, plan de mise à niveau, nouvelle revue de la porte S5 → S6

Les chiffres ci-dessus proviennent de l'étude de cas sur données synthétiques, ils illustrent la méthode et ne valent pas validation sur systèmes réels.

13. Les erreurs à éviter

Erreur	Conséquence	Correction
Publier un score moyen	Masque une exposition juridique	Publier le niveau du maillon faible
Déclarer un niveau sans dossier	Niveau indéfendable en audit	Référencer chaque preuve
Ajouter des effets indirects aux Shapley causaux	Double comptage (10,32 au lieu de 5 sur l'exemple jouet)	Utiliser $\phi C = \phi D + \psi$, $\Sigma\psi = 0$
Parler de causalité sans graphe défendable	Conclusions biaisées	Publier ϕD seule et déclarer la limite
Présenter un ECS sans ε ni témoins	Score non interprétable	Appliquer les cinq règles de publication
Lire ECS = 1 comme une bonne explication	Un explicateur constant obtient 1	Exiger la fidélité à côté de l'ECS
Utiliser les seuils RAD 0,10 et 0,25 de la v3.2	Seuils retirés car non reproductibles	Aucune lecture normative avant calibrage
Comparer deux organisations par le RAD	Comparaison prématurée	Attendre la calibration sur systèmes réels
Traiter les niveaux S comme des intervalles	Écarts mal interprétés	Compter des portes manquantes
Ne pas rétrograder après réentraînement	Système non requalifié	Appliquer S9 → S7 et comparer au jeu de référence
Ne pas mettre la fiche à jour après un incident	Traçabilité rompue	Versionner la fiche à chaque incident

14. Questions fréquentes

AEXIA est-il une certification ?

Non. C'est un cadre de gouvernance en accès ouvert. Les niveaux sont des positionnements documentés, pas des labels délivrés par un tiers.

Faut-il utiliser DCME partout ?

Non. Il est utile quand une caractéristique protégée ou un proxy peut agir par des variables intermédiaires. Ailleurs, SHAP et l'ECS suffisent.

Comment choisir la criticité β ?

Le Framework la rattache à la classification AI Act (1 à 4) sans fixer la correspondance. Repère de ce guide, à déclarer : 1 minimal, 2 limité, 3 haut risque, 4 haut risque à fort enjeu pour les personnes (crédit, emploi, justice).

Le seuil ECS ≥ 0,8 est-il obligatoire ?

Non. C'est un exemple de seuil de politique : à choisir, justifier et publier avec ε.

À quelle fréquence évaluer ?

Organisation : une à deux fois par an. Système : à chaque porte et incident. Indicateurs : chaque trimestre.

Peut-on s'en servir pour évaluer un fournisseur ?

Oui, en demandant sa fiche et ses dossiers de preuves. Sans accès aux preuves, le niveau reste « non établi ».

Quel lien avec ISO/IEC 42001 ?

La norme fournit le système de management ; AEXIA ajoute la maturité par système, les portes et les mesures d'explicabilité.

Que faire sans second évaluateur ?

Documenter l'absence et faire relire les niveaux par le DGC avant toute communication externe.

15. Feuille de route de mise en œuvre

Horizon	Mois	Priorités	Cible O	Cible S
Court terme	M0-M3	Inventaire, classification AI Act, première AEXIA Card, désignation des rôles	N0 → N1	Tous positionnés
Moyen terme	M3-M12	Politique, portes S5 → S6 et S7 → S8, registre des explications, pipelines de surveillance	N1 → N2	Aucun haut risque au-delà de S7 sans porte
Long terme 1	M12-M24	Audit interne, rétrogradation testée, automatisation, formalisation de DCME, ECS et RAD	N2 → N3	Distribution stabilisée
Long terme 2	M24-M36	Conformité annexe III avant le 2 décembre 2027, normalisation ISO, publication	N3 → N4	Rétrogradation automatisée

Limites et statut

- Validation préliminaire sur données synthétiques et petits jeux publics ; pas de validation sur systèmes réels.
- Fiabilité inter-évaluateurs des niveaux S non encore mesurée (objectif de l'étude : alpha de Krippendorff ≥ 0,70).
- Seuils (ECS, RAD, indicateurs) : repères de politique, provisoires.
- Section réglementaire datée du 22 octobre 2026 ; revue semestrielle nécessaire.
- Ce guide ne constitue pas un avis juridique.

Annexe A : Glossaire

Terme	Définition
AEXIA-O / AEXIA-S	Axes organisationnel (6 piliers, N0-N5) et système (S0-S9)
AEXIA Card	Fiche d'identité d'un système d'IA : niveau, méthode, données, surveillance, conformité
DGC	Comité de gouvernance des données et de l'IA, qui décide des portes critiques
Porte probante	Passage de niveau subordonné à un jeu fermé d'artefacts examinés en revue
Maillon faible	Règle : le niveau est celui du composant le plus faible
Rétrogradation	Baisse du niveau S après dérive, incident, changement de population, évolution réglementaire ou réentraînement
DCME	Présentation multi-échelle des valeurs de Shapley causales : $\phi C = \phi D + \psi$
ϕD, ϕC, ψ	Contribution directe, contribution causale totale, redistribution médiée (somme nulle)
ECS	Score de stabilité locale d'une explication sur un voisinage de rayon ε
Fidélité	Accord entre l'explication et le comportement du modèle (par suppression ou insertion de variables)
Témoin aléatoire / constant	Explicateurs de référence : plancher attendu proche de 0 et plafond trompeur égal à 1
RAD-lite	Part pondérée des niveaux de maturité manquants du portefeuille, entre 0 et 1
Ancienneté de l'écart	Mois écoulés depuis l'ouverture de l'écart non résorbé le plus ancien
β, S_att, S_eff	Criticité (1 à 4), niveau S attendu, niveau S effectif
TriAxis	Référentiel de biais de l'initiative : taxonomie Origine × Lifecycle × Spécifique

Annexe B : Check-lists

Avant une décision de porte

☐ Tous les artefacts de la porte sont réunis et référencés dans la fiche

☐ Le dossier a été examiné en revue par les décisionnaires prévus

☐ La classification AI Act est à jour

☐ Le maillon faible est identifié et son plan est daté

☐ La décision, la date et l'instance sont consignées

Avant de publier un ECS

☐ ε, N, normalisation

☐ Mesure de fidélité et sa valeur

☐ Témoins aléatoire et constant

☐ Plusieurs rayons et intervalle de confiance bootstrap

☐ Seuil de politique justifié

Avant de publier un DCME

☐ Graphe publié et justifié

☐ Identifiabilité défendue, sinon ϕD seule

☐ Efficacité et Σψ = 0 vérifiées

☐ Analyse de sensibilité au graphe

☐ Comparaison de ϕD à SHAP interventionnel

Avant de publier un RAD

☐ S_att et β déclarés par politique

☐ Écart lu en portes manquantes

☐ Ancienneté rapportée à part

☐ Aucune lecture normative des valeurs

☐ Si variante cumulative : T½ déclaré et justifié

Références principales

Sidibe, M. (2026). AEXIA™ Framework, version 3.3. AI BIAIS INITIATIVE. CC BY 4.0.

Lavin, A. et al. (2022). Technology readiness levels for machine learning systems. Nature Communications, 13, 6039.

Règlement (UE) 2024/1689 (règlement sur l'IA) ; règlement (UE) 2026/1744 modifiant son calendrier d'application.

NIST (2023). AI Risk Management Framework 1.0. ISO/IEC 42001:2023.

Heskes, T. et al. (2020). Causal Shapley Values. NeurIPS.

Alvarez-Melis, D., Jaakkola, T. (2018). On the Robustness of Interpretability Methods. ICML Workshop.

Stevens, S. S. (1946). On the Theory of Scales of Measurement. Science, 103.

Cunningham, W. (1992). The WyCash Portfolio Management System. OOPSLA.

Avant diffusion, vérifier chaque référence sur sa source d'origine.