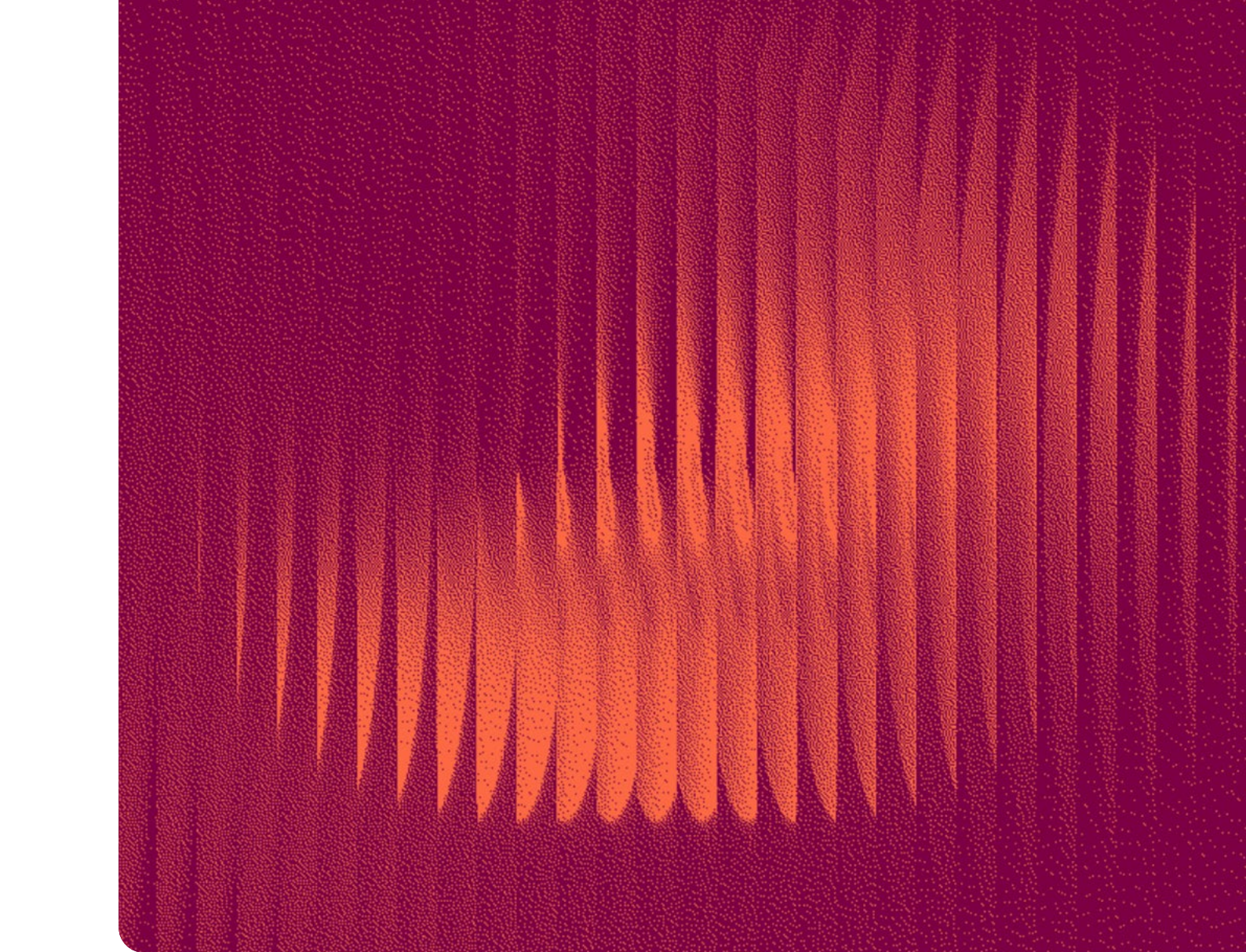




Déployer une IA conforme : pourquoi l'explicabilité n'est plus une option

Ce que le RGPD et l'AI Act imposent à vos projets d'IA



Un guide pratique pour sécuriser vos projets d'IA

L'objectif de cet ebook est d'aider les DSI, responsables data, RSSI et métiers à traduire la notion de XAI en réalité opérationnelle :

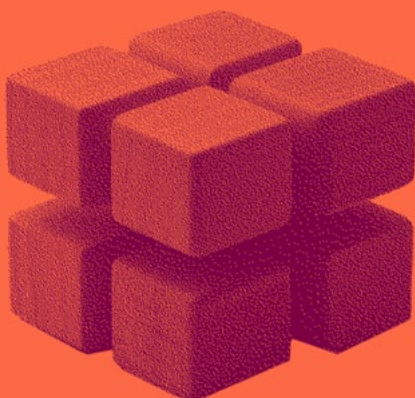
- Comprendre les enjeux réglementaires
- Clarifier ce que l'IA explicable doit permettre sur le terrain
- Identifier les capacités techniques indispensables pour rendre l'XAI opérationnelle

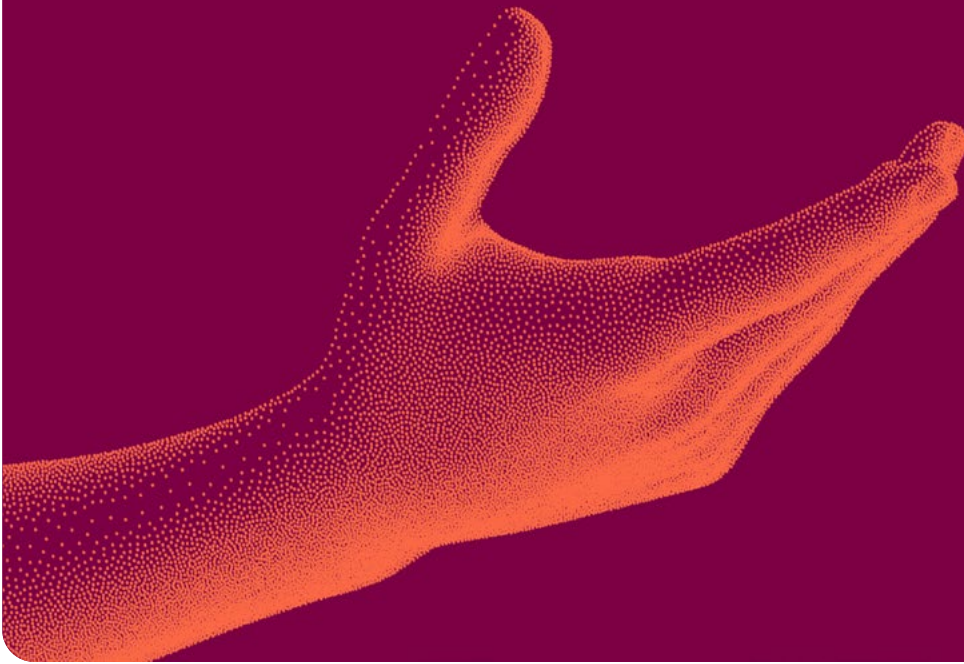
XAI ou « IA explicable » : un enjeu central pour le déploiement de l'IA en production

À mesure que les systèmes d'IA interviennent dans des décisions à fort impact humain – crédit, recrutement, diagnostic, allocation de droits – la capacité à expliquer ces décisions n'est plus une "option", mais une condition de viabilité. Une IA performante mais opaque expose l'organisation à des risques juridiques, réglementaires et réputationnels croissants.

L'IA explicable, ou Explainable AI (XAI), désigne l'ensemble des méthodes et des pratiques permettant de comprendre et d'expliquer pourquoi un modèle d'intelligence artificielle prend une décision donnée : quel raisonnement a conduit à ce résultat, quels facteurs ont influencé la décision de quelle manière ont-ils été combinés ?

Dans le cadre européen, cette exigence est désormais clairement formalisée. Le RGPD, puis l'AI Act, imposent des obligations de transparence, de traçabilité et de supervision humaine pour les systèmes d'IA, en particulier lorsqu'ils sont qualifiés de systèmes « à haut risque ». En pratique, cela signifie qu'une IA qui ne peut pas être expliquée est, à terme, une IA qui ne peut pas être durablement déployée.





01

**Le cadre normatif :
une IA non explicable
est une IA non déployable**

L'explicabilité n'est pas un choix d'architecture ou une bonne pratique facultative. Elle est devenue une **exigence réglementaire explicite**, encadrée par le droit européen applicable aux systèmes d'IA. Dès lors qu'une décision automatisée a un impact significatif sur des individus, l'organisation doit être en mesure d'en **expliquer le fonctionnement, la logique et les effets**.

Autrement dit, une IA qui ne peut pas être expliquée est, juridiquement, une IA qui expose l'organisation à un risque de non-conformité – et donc une IA difficilement déployable à grande échelle.

Le RGPD : les fondations juridiques de l'explicabilité

Le RGPD a posé les premières bases de l'obligation d'explicabilité.

Son **article 22** encadre les décisions fondées exclusivement sur un traitement automatisé produisant des effets juridiques ou des effets significatifs sur les personnes. Il reconnaît notamment le droit de ne pas faire l'objet d'une telle décision sans garanties appropriées.

Les **articles 13 à 15** complètent ce cadre en imposant une obligation d'information. Les personnes concernées doivent être informées à la fois de l'existence d'une prise de décision automatisée, et de la logique sous-jacente au traitement.

En pratique, cela signifie qu'une organisation doit être capable d'expliquer, a minima, comment et pourquoi une décision a été produite, dans un langage compréhensible pour la personne concernée. La complexité ou l'opacité d'un système d'IA ne constitue pas une justification acceptable. Autrement dit, l'explicabilité est une véritable condition d'exercice des droits des personnes.

L'éclairage de la CNIL : une exigence concrète, pas théorique

Cette exigence n'est pas abstraite. Elle a déjà donné lieu à des contrôles et à des mises en demeure.

À la suite du contrôle d'une plateforme de pré-inscription dans l'enseignement supérieur, la CNIL a constaté une absence d'information sur l'utilisation d'un algorithme et sur son rôle dans le classement et l'affectation des candidats. Cette situation a été qualifiée de manquement, car les personnes concernées ne disposaient pas des éléments leur permettant de comprendre et de contester la décision automatisée.

Dans ce cas précis, l'autorité a demandé de faire cesser les décisions fondées exclusivement sur un traitement automatisé, et d'introduire une intervention humaine effective, capable de tenir compte des observations des personnes concernées.

Ce point est essentiel : sans explicabilité, la supervision humaine exigée par les textes devient impossible à exercer en pratique.

L'AI Act : l'explicabilité comme pilier des systèmes à haut risque

L'AI Act européen vient renforcer et structurer ces exigences, en particulier pour les systèmes d'IA dits « à haut risque. »

- L'**article 13** impose des obligations de transparence et d'information visant à permettre aux utilisateurs de comprendre le fonctionnement du système et d'interpréter correctement ses résultats.
- L'**article 14** complète cette approche en exigeant une supervision humaine effective, ce qui suppose que les décisions de l'IA soient compréhensibles, explicables et contestables.

Sans mécanismes d'explicabilité, ces obligations restent purement théoriques. Il est impossible de superviser, corriger ou encadrer un système dont les décisions ne peuvent pas être expliquées.



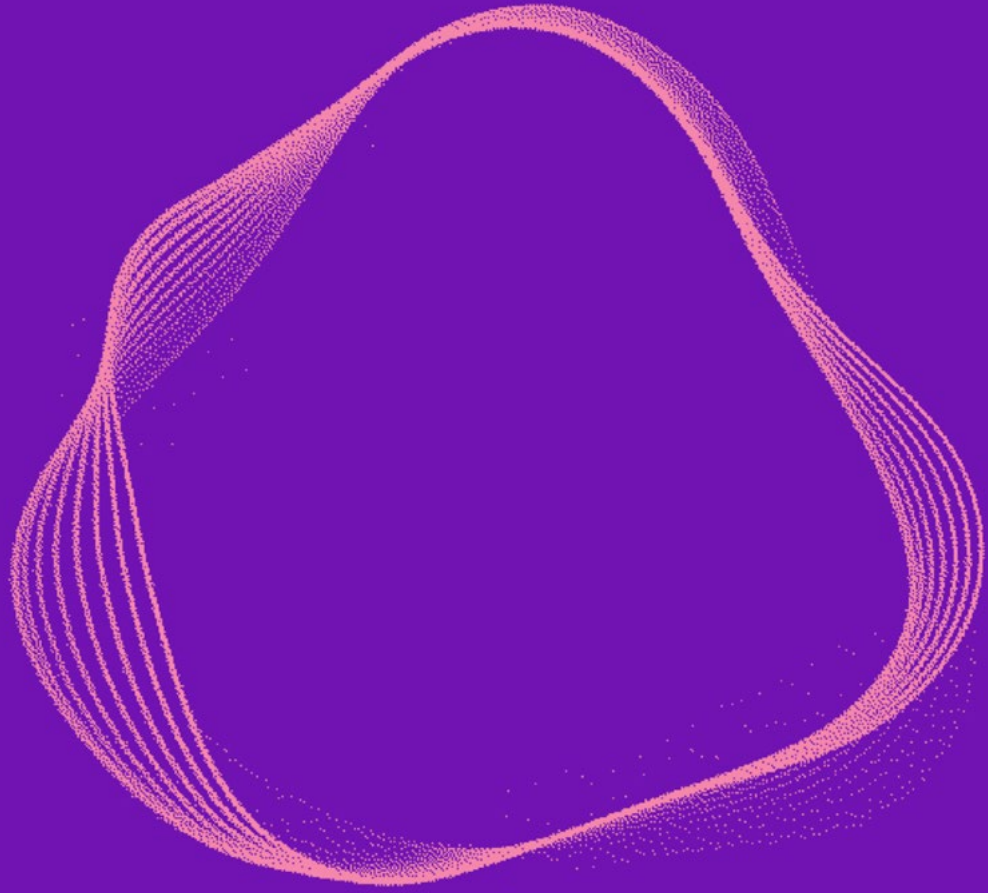
Des conséquences très concrètes

Pour les équipes IT et data, ces exigences changent profondément la nature des projets IA. Il ne s'agit plus seulement de livrer un modèle performant, mais de garantir :

- la traçabilité des décisions dans le temps,
- la conservation des explications associées,
- et la preuve de l'intégrité des résultats, c'est-à-dire l'assurance qu'ils n'ont pas été altérés.

En cas de contrôle, de litige ou de demande d'un régulateur, l'organisation doit pouvoir démontrer que chaque décision automatisée repose sur un processus maîtrisé, documenté et reproductible.

Dans ce contexte, l'XAI devient un pilier de la conformité réglementaire, en particulier dans les secteurs sensibles comme la finance, la santé, les ressources humaines ou la justice. Pour les organisations qui déploient leurs modèles sur des infrastructures souveraines et de confiance, l'explicabilité est indissociable de la gouvernance de l'IA : elle conditionne à la fois la conformité aux textes européens et la confiance des utilisateurs finaux.



02

**Initier une démarche d'IA
explicable : quelle réponse
opérationnelle ?**

Concrètement, l'IA explicable ne se résume ni à un outil, ni à une obligation documentaire.

La démarche suppose de structurer les projets d'IA dès l'amont pour répondre à des exigences claires. Les 4 points suivants constituent les fondations d'une XAI réellement opérationnelle.

01

Rendre les décisions compréhensibles et exploitables par les métiers

Le premier objectif de l'IA explicable est la compréhension métier. Il ne s'agit pas d'expliquer un algorithme ou une formule mathématique, mais de rendre une décision intelligible : pourquoi ce dossier a été refusé, pourquoi cette alerte a été déclenchée, ou pourquoi cette recommandation a été formulée.

Pour être réellement utile, cette compréhension doit être exploitable. Une explication n'a de valeur que si elle permet d'agir, de vérifier ou de discuter une décision. Cela suppose de dépasser les scores ou indicateurs bruts pour mettre en évidence les facteurs réellement déterminants : quels critères ont pesé le plus, dans quel sens, et avec quel poids relatif.

Cette explicabilité doit également être **adaptée aux profils concernés**.

Les équipes opérationnelles attendent des explications simples et actionnables. Les managers cherchent à comprendre des logiques globales pour piloter l'activité. Les fonctions de contrôle, enfin, ont besoin d'éléments plus détaillés et traçables. Une XAI efficace est donc capable de proposer plusieurs niveaux de lecture, sans complexifier l'usage.

Sans cette capacité, même un modèle très performant reste perçu comme une boîte noire, et risque d'être ignoré ou rejeté par les équipes.



02

Garantir l’auditabilité et la traçabilité des décisions

Au-delà de l’usage métier, l’XAI doit permettre l’auditabilité des décisions. Une organisation doit être en mesure de revenir sur une décision passée et de répondre à des questions structurantes : quel modèle a été utilisé, avec quelles données, selon quelle version, et avec quels éléments explicatifs associés.

Cette capacité est indispensable dans les secteurs réglementés, mais aussi pour la gestion des incidents, des litiges ou des contrôles internes. Elle permet de démontrer que les décisions automatisées reposent sur un processus maîtrisé, documenté et reproductible.

Sans traçabilité, l’IA devient un point aveugle du système d’information : difficile à expliquer, impossible à défendre et risquée à exploiter.

03

Rendre les décisions contestables et améliorables dans le temps

Une décision automatisée ne doit jamais être perçue comme définitive ou incontestable par nature. Elle doit pouvoir être questionnée, expliquée, documentée et, si nécessaire, corrigée.

L'XAI fournit les éléments factuels visant à justifier une décision, en expliciter les limites et ouvrir des mécanismes de correction. Elle permet aux utilisateurs de signaler une recommandation incorrecte, inapplicable ou inadaptée au contexte réel.

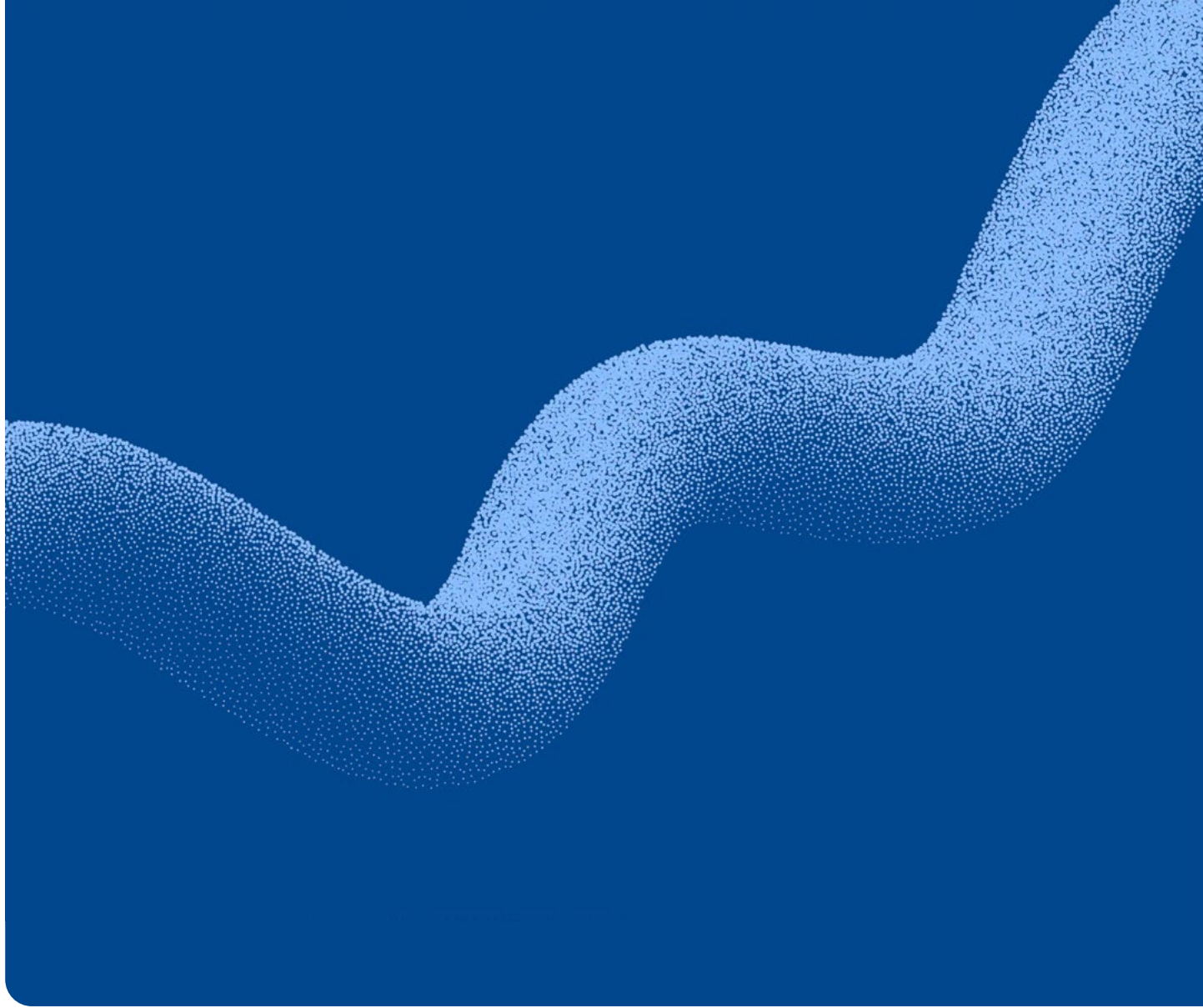
Ces retours ne servent pas uniquement à corriger des cas isolés. Ils alimentent une logique d'amélioration continue, dans laquelle l'IA progresse grâce à l'expérience humaine. L'explicabilité devient alors un levier d'alignement durable entre le modèle et la réalité opérationnelle.

04

Détecter et prévenir les biais

Enfin, l'explicabilité joue un rôle central dans la détection des biais. En rendant visibles les corrélations sur lesquelles s'appuie un modèle, elle permet d'identifier des logiques discriminantes ou des effets indésirables issus des données d'entraînement.

Cette transparence est indispensable pour prévenir les dérives, limiter les risques juridiques et répondre aux exigences croissantes en matière d'IA responsable. Elle contribue également à renforcer la confiance des utilisateurs finaux, en montrant que les décisions automatisées reposent sur des critères identifiables, maîtrisés et discutables.



03

Les fondations techniques d'une IA explicable et conforme



L'IA explicable ne se joue pas uniquement au niveau du modèle. Sa valeur réelle dépend de l'environnement technique dans lequel les décisions sont produites et expliquées.

Pour qu'une explication soit crédible, il faut pouvoir garantir trois choses simples :

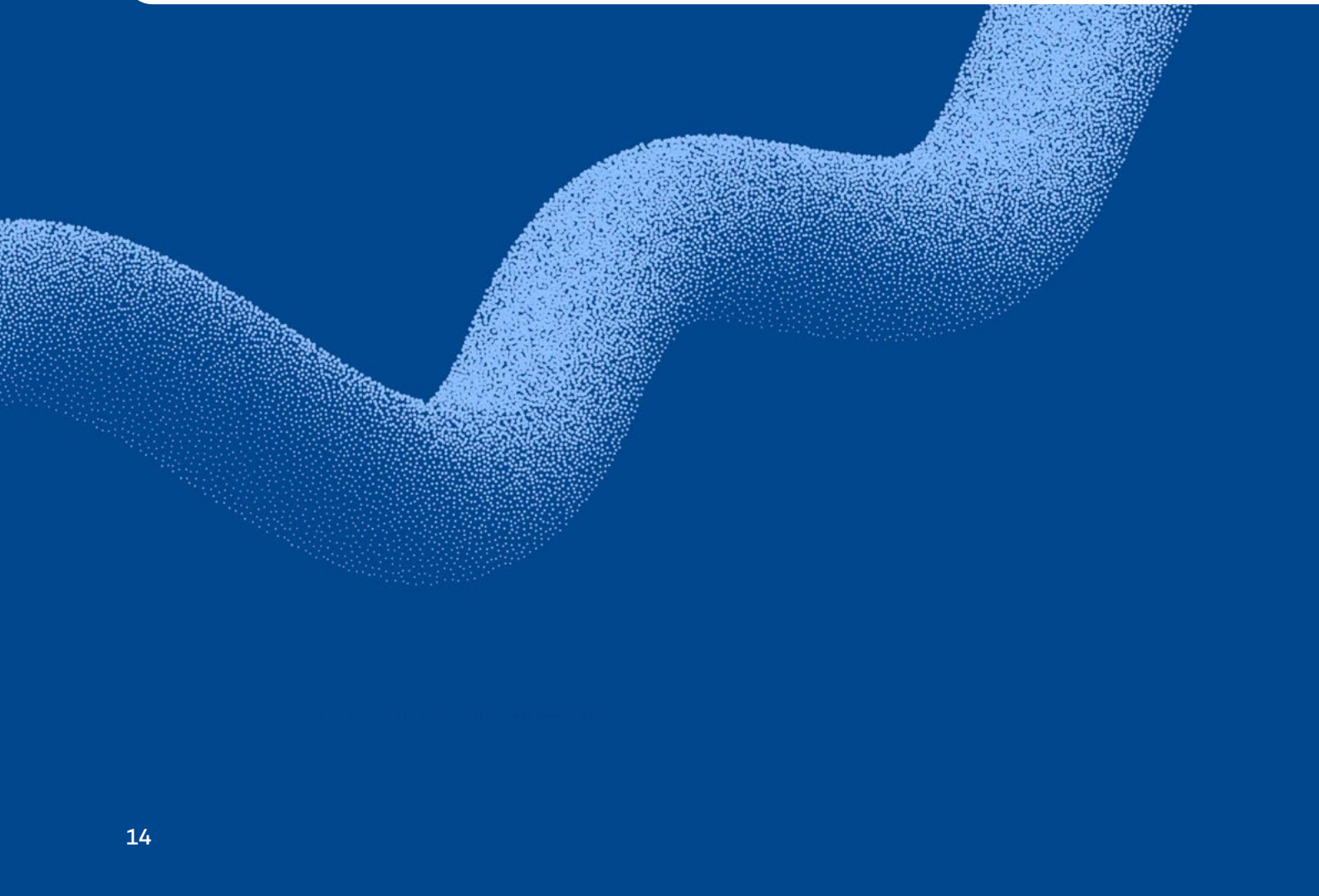
- que le modèle qui prend la décision est bien celui qui a été validé,
- que l'explication correspond exactement à la décision produite,
- et que l'ensemble n'a pas été modifié, ni techniquement ni humainement.

Sans cela, l'explicabilité reste théorique. Elle existe peut-être sur le plan technique, mais elle n'est ni auditable, ni juridiquement défendable.

Une plateforme adaptée doit donc permettre de répondre, à tout moment, à des questions essentielles : quel modèle a été utilisé, avec quelles données, à quelle date, et avec quelle justification associée. Elle doit également produire un journal d'audit fiable et opposable, capable d'être présenté en cas de contrôle, de litige ou de demande d'un régulateur.

Dans les secteurs régulés – santé, finance, secteur public, justice – cette exigence est encore plus forte. Il ne suffit pas d'avoir une IA « qui fonctionne ». Il faut une IA hébergée dans un environnement de confiance, conforme aux exigences européennes (RGPD, AI Act), et auditable à tout moment. Une explication n'a de valeur juridique que si elle est infalsifiable.

Cela implique une maîtrise complète de la chaîne de traitement : savoir où sont les données, qui y accède, comment elles sont utilisées, et comment les décisions sont produites et expliquées. Dès lors qu'une décision impacte un patient, un client, un citoyen ou un recrutement, la question n'est plus seulement technique, elle devient juridique et stratégique.



**Une plateforme XAI-ready
ne se résume pas à un tableau de bord ;
c'est un écosystème de confiance
capable de transformer une « boîte noire »
en une « boîte de verre » auditable, souveraine
et juridiquement opposable.**

Aâdel Benyoussef, Strategic Accounts & GTM IA/Data Lead chez Numspot



Checklist.

Les 6 piliers d'un environnement « XAI-ready » industriel



Pilier 01

Sécurité et Souveraineté « by Design »

- **Caractéristiques**

La plateforme doit garantir une étanchéité totale. Au-delà du chiffrement, elle doit reposer sur une infrastructure déconnectée des lois extra-territoriales.

- **Impact**

Garantir que les explications du modèle (souvent révélatrices de secrets industriels ou de données sensibles) restent sous votre contrôle exclusif.

- **Critère concret**

Isolation physique/logique des environnements de calcul et stockage des données sur le territoire national.



Pilier 02

Traçabilité Native et Immuable

- **Caractéristiques**

Chaque inférence doit être liée à un « cliché » (snapshot) exact de l'instant T.

- **Impact**

Éliminer le risque d'incohérence entre le modèle audité et le modèle en production.

- **Critère concret**

Un registre centralisé et immuable (versioning automatique) liant : [Données d'entrée] + [Version du Modèle] + [Code d'Explication] + [Horodatage].



Pilier 03

Auditabilité Opérationnelle « Zero-Effort »

- **Caractéristiques**

L'explicabilité ne doit pas être un projet de recherche à chaque contrôle, mais une fonction d'exportation standard.

- **Impact**

Répondre aux régulateurs (CNIL, ACPR, AI Act) en quelques clics plutôt qu'en quelques semaines.

- **Critère concret**

Génération de rapports d'explicabilité (PDF/JSON) directement exploitables par des non-experts ou des auditeurs externes.



Pilier 04

Agnosticité et Interopérabilité (Open XAI)

- **Caractéristiques**

Éviter l'enfermement propriétaire. L'environnement doit être une «coque» accueillante pour les meilleures bibliothèques open-source.

- **Impact**

Pérenniser vos investissements en intégrant les dernières avancées (SHAP, LIME, Integrated Gradients) sans dépendre du bon vouloir d'un éditeur.

- **Critère concret**

Compatibilité totale avec les frameworks standards (Scikit-learn, PyTorch, TensorFlow) et les conteneurs (K8S/Docker).



Pilier 05

Gouvernance et Cycle de Vie Maîtrisé

- **Caractéristiques**

La confiance repose sur le « Qui fait quoi ? ». L'XAI sans gestion des droits est un risque de sécurité.

- **Impact**

Prévenir les dérives de modèles (drift) et s'assurer que seules les explications validées par le « Risk Officer » sont diffusées.

- **Critère concret**

Workflow d'approbation intégré (Human-in-the-loop) avant toute mise en production ou modification de la logique d'explication.



Pilier 06

Scalabilité Critique (Haute Disponibilité)

- **Caractéristiques**

L'explicabilité est gourmande en calcul. La plateforme doit absorber cette charge sans ralentir les décisions métier.

- **Impact**

Passer de 10 à 10 000 explications par seconde sans dégradation de service ni compromis sur la précision des calculs d'importance de variables.

- **Critère concret**

Auto-scaling des ressources de calcul sur une infrastructure cloud hybride, robuste et hautement disponible.

À propos de Numspot

Numspot est votre partenaire de confiance pour accélérer la transformation numérique de votre organisation, garantir la souveraineté de vos données et renforcer votre performance, grâce à sa plateforme qui associe l'innovation du cloud natif et l'IA.

