



L'IA en santé n'est pas uniquement un problème d'algorithme. C'est avant tout un problème d'architecture.

13 février 2026

Six dimensions pour évaluer ce que presque personne ne mesure.

Résumé

L'IA en santé échoue rarement par manque de performance statistique. Elle échoue par fragilité architecturale : Un modèle avec une AUC élevée ne garantit ni adoption clinique, ni soutenabilité organisationnelle, ni résilience opérationnelle.

La viabilité d'un déploiement dépend de six dimensions : topologie décisionnelle, bilan temporel et ROI multi-échelles, réduction d'incertitude, degré de couplage architectural, charge écosystémique et TCO, réversibilité et dégradation gracieuse.

La performance n'est pas un invariant : elle dépend d'un protocole de mesure et d'un benchmark souvent non partagé et donc "subjective". La question centrale n'est donc pas "le modèle est-il performant ?" mais "le système peut-il l'absorber sans se fragiliser ?"

Ce cadre propose une grille d'éligibilité pour distinguer les déploiements structurellement fragiles des expérimentations soutenables.

Introduction

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

Dans un article précédent ([L'IA réduit-elle la charge cognitive des cliniciens — ou l'augmente-t-elle ? | LinkedIn](#)), je posais une thèse simple : le critère déterminant de l'adoption d'une IA en contexte clinique n'est pas sa performance algorithmique isolée. C'est le ratio entre valeur clinique perçue et friction systémique induite.

Cette formulation était volontairement provocatrice. Elle mérite d'être précisée.

La performance algorithmique est une condition nécessaire. Elle n'est pas une condition suffisante. Ce qui sépare un modèle publié dans un journal d'un outil réellement intégré dans une pratique clinique, c'est la viabilité systémique de son insertion. L'écart entre la publication et la production est le cimetière silencieux de la majorité des projets d'IA en santé.

Le cadre qui suit propose une grille d'analyse en six dimensions. Il ne remplace pas l'évaluation statistique des modèles. Il ajoute un second axe, orthogonal : l'architecture du déploiement.

L'inversion nécessaire

La question prédominante que ce soit dans la littérature, dans les congrès, dans les appels à projets, reste : *comment améliorer la performance du modèle ?*

Cette focalisation, légitime à un stade technologique initial, devient problématique lorsqu'elle retarde l'examen des conditions architecturales de déploiement. Elle traduit une attention encore majoritairement portée sur l'optimisation du composant technique plutôt que sur la soutenabilité systémique de son intégration.

Mais avant même de déplacer l'unité d'analyse vers l'architecture, un point doit être explicité : la performance d'un modèle n'est pas un invariant objectif. **Elle est le produit d'un protocole de mesure.**

Elle dépend du benchmark retenu (monocentrique ou multicentrique, rétrospectif ou prospectif, nettoyé ou représentatif du bruit clinique réel). Un modèle peut atteindre une AUC de 0,95 sur un dataset interne soigneusement préparé et voir sa performance diminuer significativement en conditions réelles. De nombreux benchmarks excluent précisément ce qui rend la clinique complexe : comorbidités, données manquantes, hétérogénéité des pratiques.

Dans certaines niches d'IA clinique, les contraintes structurelles (rareté des données, régimes réglementaires, forte dimensionnalité avec faible effectif -HDLSS-, données propriétaires) ne permettent pas l'émergence de benchmarks partagés. À l'inverse, dans les grands modèles de langage, la concurrence ouverte a favorisé l'apparition de standards d'évaluation progressivement stabilisés. La performance devient alors moins

Jérôme Vetillard

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

une propriété intrinsèque qu'un attribut relatif au cadre d'évaluation choisi. En l'absence de protocoles partagés, reproductibles et validés collectivement, la comparabilité elle-même devient fragile.

Cependant, même en supposant une mesure rigoureuse, indépendante et cliniquement représentative de la "performance", une question demeure : **dans quelle architecture cette performance peut-elle produire un effet net positif ?**

Le changement d'unité d'analyse est décisif. On ne part plus de l'algorithme pour chercher son point d'insertion. On part du système socio-technique réel, multi-échelles (le service, l'établissement, le territoire) pour identifier où une capacité algorithmique peut réduire une incertitude spécifique **sans générer de complexité supérieure au gain** :

- Le même modèle, inséré dans deux architectures différentes, produira des effets radicalement différents.
- Non parce que l'algorithme change. Parce que le système dans lequel il opère change.

Cela suppose, en amont de toute évaluation, une immersion structurée dans le travail réel. Les disciplines existent : anthropologie médicale, ergonomie cognitive, analyse de l'activité, shadowing clinique. Elles sont rarement mobilisées au stade de la conception. C'est pourtant là qu'elles sont décisives.

Dimension 1 : Topologie décisionnelle

Avant toute considération algorithmique (valable également pour les solutions logicielles "classiques" hors du champ de l'intelligence artificielle, et là où ces solutions pèchent depuis des décennies), il serait nécessaire de cartographier la structure **réelle** de la décision.

Une IA ne s'insère pas dans "la décision médicale" en général. Elle s'insère dans une topologie décisionnelle spécifique. Quatre configurations minimales peuvent être distinguées :

- Décision individuelle synchrone (urgence, réanimation).
- Décision individuelle asynchrone (consultation programmée).
- Décision collégiale synchrone (RCP, staff).
- Décision distribuée asynchrone (parcours chronique multi-acteurs).

Chaque contexte peut être décrit par un triplet minimal :

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

- **D** : durée moyenne disponible par cas.
- **N** : nombre d'acteurs impliqués.
- **L** : tolérance maximale à la latence d'information.

Prenons deux exemples concrets.

En **Réunion de Concertation Pluridisciplinaire (RCP) oncologique** : $D \approx 5$ minutes par dossier, $N \approx 6$ à 10 praticiens (oncologue, chirurgien, radiothérapeute, anatomopathologiste, radiologue...), L proche de zéro — toute information doit être immédiatement disponible dans l'espace de discussion. Un outil qui nécessite trois minutes d'explication dans ce créneau de cinq minutes ne sera pas adopté. L'incompatibilité n'est pas algorithmique. Elle est topologique.

En **consultation de médecine générale** : $D \approx 15$ à 20 minutes, $N = 1$, L modérée (le médecin peut consulter un résultat, réfléchir, y revenir). La même recommandation algorithmique, inacceptable en RCP, peut ici s'intégrer sans friction.

Toute évaluation sérieuse devrait expliciter le triplet (D , N , L) du contexte cible. Sans cela, la performance du modèle reste abstraite, "hors sol".

Dimension 2 : Bilan temporel net et retour sur investissement multidimensionnel

Le clinicien effectue déjà, implicitement, à chaque interaction avec un outil numérique, un calcul simple :

$$T_{\text{net}} = T_{\text{restitué}} - (T_{\text{apprentissage}} + T_{\text{intégration}} + T_{\text{évaluation}} + T_{\text{coordination}})$$

- **T_apprentissage** : appropriation initiale et réapprentissage liés aux évolutions du modèle. Un modèle qui évolue impose un coût récurrent.
- **T_intégration** : login, navigation, bascule d'écran, saisie complémentaire. Ce temps est souvent invisible dans les évaluations mais cumulativement massif. Quelques secondes par patient deviennent des heures à l'échelle d'un service.
- **T_évaluation** : vérification, interprétation et mise en contexte de la sortie algorithmique. Plus le modèle est opaque, plus ce temps augmente. L'explicabilité est un déterminant opérationnel.
- **T_coordination** : discussions supplémentaires induites — arbitrage en RCP, justification d'une recommandation divergente, ajustement collectif du niveau de confiance accordé au modèle.

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

Si T_net est durablement négatif, l'outil ne sera pas intégré, quelle que soit sa performance statistique.

Ce critère est falsifiable. Il peut être mesuré en conditions réelles d'usage, interruptions comprises. Il doit être évalué longitudinalement : un T_net initialement négatif peut devenir positif avec l'appropriation. À l'inverse, il peut se dégrader si le modèle dérive, si les données d'entrée évoluent ou si l'organisation change.

Ce bilan temporel ne concerne pas que le clinicien. Le temps perdu ou gagné par la DSI, par la gouvernance, par l'organisation est tout aussi réel. C'est l'objet de la dimension 5.

Du bilan temporel au retour sur investissement

En médecine de ville, le lien entre gain de temps et impact économique est relativement direct : le décideur, l'utilisateur et le bénéficiaire sont souvent la même personne.

À l'hôpital, l'équation change de nature. Le ROI dépend de l'échelle à laquelle on l'observe :

- **Niveau clinicien** : gain de temps, réduction de charge cognitive, diminution de l'incertitude mais qui ne se traduit pas automatiquement en économie budgétaire.
- **Niveau service** : fluidification des RCP, diminution d'examens redondants, optimisation de la programmation mais le bénéfice se traduit souvent en amélioration des flux plutôt qu'en réduction de masse salariale.
- **Niveau établissement** : réduction de la durée moyenne de séjour, diminution des complications évitables, amélioration des indicateurs qualité mais avec des effets souvent indirects, différés, et dépendants du modèle de financement.
- **Niveau territorial ou payeur** : prévention d'événements graves, meilleure allocation des ressources mais le ROI peut être positif pour l'assurance maladie et neutre, voire négatif, pour l'établissement.

Ces niveaux ne convergent pas nécessairement. Celui qui investit n'est pas toujours celui qui bénéficie. Le coût est immédiat, le bénéfice souvent différé. Un outil peut générer un ROI positif à l'échelle micro tout en créant une charge systémique supérieure à l'échelle macro.

Ce décalage ne peut être résolu indépendamment de l'architecture : un couplage fort augmente les coûts de maintenance et de PRA, une dépendance contractuelle limite la capacité de négociation, une absence de réversibilité transforme un investissement

Jérôme Vetillard

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

initial en contrainte structurelle. Un ROI financier apparent ne compense pas une fragilité architecturale.

Le bilan temporel et le ROI ne sont donc pas des objectifs en soi. Ce sont des indicateurs de viabilité dans un système plus large. Et le temps, même correctement valorisé, n'est pas la seule monnaie pertinente.

Dimension 3 : Réduction d'incertitude décisionnelle

Un chirurgien hésite entre deux stratégies opératoires. Un oncologue est confronté à un profil atypique en deuxième ligne de traitement. Un radiologue revoit une image ambiguë. Dans ces situations, la valeur clinique ne se mesure pas en minutes gagnées. Elle se mesure en **incertitude réduite**.

On peut introduire une métrique complémentaire le delta d'incertitude :

$\Delta I = I_{\text{avant}} - I_{\text{après}}$ où I peut être approché par :

- variance inter-cliniciens sur une décision donnée,
- taux de discordance en RCP,
- fréquence de requalification diagnostique,
- variabilité des prescriptions à présentation clinique comparable.

Ces mesures existent dans la littérature mais sont rarement intégrées dans les protocoles d'évaluation en conditions réelles. Ce qui est précisément le problème. Un outil peut augmenter légèrement le temps de décision mais être accepté (et même recherché) si la réduction d'incertitude est substantielle et mesurable. Le radiologue qui passe trente secondes de plus sur une image mais tranche avec plus de sûreté clinique un diagnostic différentiel ne considère pas ces trente secondes comme une perte.

Le cadre devient alors bidimensionnel : **temps et entropie décisionnelle**. Une IA est pertinente si elle améliore l'un ou l'autre sans dégrader excessivement l'autre.

Dimension 4 : Couplage architectural

Comment le composant IA s'articule-t-il avec le système d'information existant ?

Le couplage ne décrit pas seulement une intégration technique. Il décrit le degré d'interdépendance structurelle entre le composant IA et le reste du SI. On peut définir trois configurations types :

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

- **Encapsulé (couplage faible).** Le composant est autonome, interfacé via API ou événements, versionnable indépendamment, désactivable sans altération majeure du fonctionnement global. L'intelligence est insérée à un point précis de la chaîne de valeur, dans un périmètre fonctionnel clairement délimité. La substitution ou le retrait restent techniquement et organisationnellement maîtrisables. Par exemple, un algorithme d'aide à la décision clinique produit par TweenMe, encapsulé sous forme d'API, qui s'insère dans une chaîne applicative disposant idéalement d'un mode dégradé sans IA.
- **Intégré (couplage modéré).** L'IA est embarquée dans une application existante — par exemple une aide à la prescription au sein du DPI. Le composant dépend du cycle de vie de l'application hôte : ses mises à jour, son audit et son monitoring sont contraints par la feuille de route et la gouvernance de l'éditeur. Le retrait est possible, mais conditionné par l'architecture applicative existante.
- **Diffus (couplage fort).** L'IA est distribuée fonctionnellement dans plusieurs composants du SI. Son périmètre peut être clairement défini, mais il traverse plusieurs couches applicatives et dépendances techniques. On évoque parfois la notion d'intelligence "ambiante" (diffuse et omniprésente). L'audit et le versionning restent possibles, à condition d'une architecture d'observabilité et de gouvernance mature. En revanche, la complexité systémique augmente : multiplication des interfaces, coordination inter-équipes accrue, cycles de version désynchronisés. Le coût de mise à jour, de retrait ou de substitution devient plus élevé. **PREDICARE ([PREDICARE : Passer d'une médecine de l'aigu à une médecine non seulement préventive, mais surtout prédictive | LinkedIn](#))** est un exemple d'architecture à couplage fort assumé (car c'est une architecture nativement agentique) : le jumeau numérique territorial fait appel à plusieurs algorithmes agentiques qui coopèrent pour fournir le service d'ensemble.

Mais le couplage doit être évalué sur **trois plans**, pas un seul :

- **Technique** : dépendances logicielles réelles, architecture d'intégration, middleware, gestion des versions.
- **Contractuel** : accès aux poids du modèle, aux logs d'inférence, aux paramètres critiques, conditions de résiliation, portabilité des données et des artefacts dérivés.
- **Réglementaire** : capacité de l'établissement à produire la documentation exigée (traçabilité, gestion du cycle de vie, surveillance post-market) sans dépendre

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

exclusivement du fournisseur et imposée notamment par l'IA Act qui classe les algorithmes d'intelligence artificielle dans la catégorie à "Haut Risque" et qui définit donc de nombreuses contraintes d'implémentation.

Un composant techniquement encapsulé peut rester fortement couplé sur le plan contractuel ou réglementaire. La maîtrise du code ne garantit pas la maîtrise stratégique. C'est une situation que plusieurs DSI découvrent après signature, quand le couplage contractuel révèle ce que l'architecture technique dissimulait.

Un composant intégré peut créer de telles dépendances techniques que les coûts de changement du composant (par exemple changer d'éditeur pour une solution de gestion des dossiers patients) peuvent être perçus comme rédhibitoires (notamment du fait des risques opérationnels) créant, ou visibilisant le "lock-in".

D'où une règle de conception : **le degré de couplage doit être proportionnel à la maturité de preuve du modèle et à la criticité de la décision supportée.**

Un modèle en phase exploratoire n'a aucune raison d'être en couplage fort avec un DPI. À l'inverse, un modèle stabilisé, validé et gouverné peut justifier un niveau d'intégration plus profond à condition que la réversibilité reste pensée et anticipée architecturalement.

Dimension 5 : Charge écosystémique et coût total de possession

L'erreur classique : évaluer l'impact de l'IA uniquement du point de vue de l'utilisateur final. Le clinicien gagne cinq minutes. Personne ne demande ce que la DSI a perdu pour les lui offrir ou le service achat pour la lui fournir.

Mais l'erreur classique, plus profonde encore est de ne pas penser *A Priori* les moyens de la mesure de l'efficacité médico-économique de la solution.

L'IA génère une charge distribuée sur l'ensemble de l'écosystème :

- **Charge clinicien** : cognitive, temporelle, attentionnelle (couverte par la dimension 2).
- **Charge DSI** : intégration, maintenance, monitoring, cybersécurité, gestion des montées de version, coordination multi-éditeurs, surveillance des dérives.
- **Charge gouvernance** : conformité réglementaire, documentation technique, surveillance post-market, gestion des incidents, préparation aux audits.
- **Charge organisationnelle** : formation initiale et continue, conduite du changement, ajustement des processus.

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

- **Charge souveraineté** : choix d'infrastructure (on-premise vs cloud souverain vs hybride), localisation des données, dépendance stratégique, exposition aux régimes juridiques étrangers.

À cela s'ajoute le **coût total de possession** (TCO), rarement calculé en amont :

- coûts d'intégration initiale,
- coûts de maintenance continue (modèle, données, infrastructure),
- coûts réglementaires (documentation, audits, surveillance post-market),
- coûts de réversibilité (que coûte le retrait de l'outil si nécessaire ?).

Un outil peut avoir un T_net positif localement et rester globalement non viable si la charge DSI ou réglementaire excède la capacité organisationnelle.

C'est exactement le piège identifié dans le chapitre "trente ans d'empilement numérique hospitalier" de l'article précédent : chaque couche semblait résoudre un problème local tout en déplaçant la complexité vers d'autres acteurs du système :

- Le DPI a soulagé certains flux mais il a transféré du temps clinique vers la documentation.
- La prescription informatisée a réduit certaines erreurs mais elle a créé de nouvelles frictions d'interopérabilité.

Les établissements disposent d'un **budget implicite de traitement de la complexité**. Chaque couche supplémentaire consomme ce budget. L'IA n'est pas exemptée de cette comptabilité. Chaque nouveau projet devrait pouvoir réduire *In Fine* cette complexité (réduire le coût marginal de modification du Système d'Information -SI-), et augmenter la capacité d'adaptation du SI à de nouveaux enjeux.

Dimension 6 : Réversibilité et dégradation gracieuse

Un système clinique est un système critique. L'introduction d'un composant IA ne doit jamais transformer une assistance (clinique, bureautique...) nouvelle en point de défaillance unique.

La persistance du support papier dans les services hospitaliers n'est pas un archaïsme, c'est l'illustration quotidienne d'une réversibilité qui fonctionne. Si votre service ne peut pas basculer en mode fiche bristol le jour où l'IA tombe, vous avez un problème de réversibilité et de capacité à fonctionner en mode dégradé.

Trois exigences doivent être distinguées :

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

Réversibilité fonctionnelle. Le système doit pouvoir fonctionner sans le composant IA. Sa désactivation (maintenance planifiée, défaillance, cyberattaque, perte de confiance clinique) ne doit pas provoquer d'effondrement fonctionnel. Un plan de retrait doit être formalisé et testé en simulation, pas simplement déclaré dans un document qualité que personne ne relit.

Si un hôpital ne peut plus fonctionner correctement le jour où l'IA est indisponible, c'est que l'intégration a créé une **dépendance structurelle non maîtrisée**.

Redondance et continuité d'activité. La résilience ne se limite pas à la possibilité de retrait. Elle implique :

- redondance d'infrastructure (multi-nœuds, multi-zones, sauvegardes testées),
- séparation des plans de données et de décision lorsque nécessaire,
- existence d'un Plan de Reprise d'Activité (PRA) et d'un Plan de Continuité d'Activité (PCA) intégrant explicitement le composant IA,
- délais de bascule compatibles avec la criticité clinique de la décision supportée.

Dans les établissements classés opérateurs d'importance vitale, cette exigence inclut l'analyse des dépendances d'infrastructure et des contraintes juridiques associées (y compris des réflexions de souveraineté numérique). La continuité ne peut dépendre d'un acteur unique sans mécanisme de repli maîtrisé.

Dégradation gracieuse. La réversibilité traite de l'indisponibilité. La dégradation gracieuse traite de la performance dégradée qui est un problème plus insidieux.

En cas de "data drift" ou de modification de la population :

- indicateurs surveillés en continu,
- seuils explicites de déclenchement d'alerte,
- réduction progressive du périmètre d'intervention,
- interdiction absolue de sortie silencieuse non signalée.

Un modèle indisponible est visible. Un modèle dégradé silencieusement est dangereux. Et pas seulement pour le patient concerné : un modèle qui trompe silencieusement érode la confiance non seulement dans cet outil, mais dans la catégorie entière des outils d'IA clinique.

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

La dégradation gracieuse doit être testée périodiquement, comme un exercice incendie : simulation de données dégradées, vérification du comportement du système, mesure du temps de bascule vers le fonctionnement sans IA.

Comment ces dimensions interagissent

Ces six dimensions ne sont pas indépendantes. Elles forment un système de contraintes couplées.

Le **couplage architectural** (4) influence directement la **charge écosystémique** (5), la **réversibilité et la redondance** (6), et la complexité des plans de continuité et de reprise d'activité. Plus le couplage est fort et diffus, plus la coordination inter-équipes est exigeante, plus le PRA devient complexe, et plus la dépendance systémique augmente.

La **topologie décisionnelle** (1) contraint le **bilan temporel** (2) acceptable et le niveau de résilience requis. Une RCP avec un D de cinq minutes et un N élevé tolère moins de friction, mais aussi moins d'indisponibilité. La criticité décisionnelle doit guider le niveau de redondance, la latence de la réponse du système et les délais de bascule admissibles.

La **réduction d'incertitude** (3) peut justifier une friction temporelle modérée ou un niveau d'intégration plus élevé à condition que le gain soit démontré et que la résilience soit proportionnée à la criticité de la décision.

Le **TCO et la charge écosystémique** (5) conditionnent la soutenabilité de la redondance. Un système techniquement résilient mais économiquement non soutenable reste fragile à moyen terme.

Un déficit sévère sur une seule dimension suffit à compromettre l'ensemble : une excellente performance algorithmique ne compense ni un T_{net} négatif, ni une dépendance systémique non maîtrisée, ni un PRA inexistant.

On peut en déduire une **matrice d'éligibilité minimale**.

Zone de fragilité structurelle :

- T_{net} négatif dans la topologie cible.
- Couplage fort avec modèle à maturité faible.
- Absence de PRA intégrant explicitement le composant IA.
- Réversibilité non testée en conditions simulées.
- Charge DSI ou réglementaire non compensée budgétairement.

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

Zone d'expérimentation acceptable :

- Couplage faible ou encapsulé.
- Topologie individuelle ou asynchrone.
- Monitoring actif du drift et seuils explicites.
- Réversibilité et redondance testées périodiquement.
- Impact budgétaire maîtrisé et explicitement assumé.

La plupart des projets qui échouent en production se situaient, sans le savoir, en zone de fragilité structurelle. Non par défaut de performance algorithmique. Par sous-estimation des dépendances systémiques et des exigences de résilience.

À quoi sert ce cadre concrètement

Outil de conception. Avant le développement : identifier où insérer l'IA, avec quel degré de couplage, pour réduire quelle incertitude dans quelle topologie décisionnelle et avec quel niveau de résilience requis. Cela suppose une immersion préalable dans le travail réel : pas des entretiens déclaratifs et des parcours patient modélisés sur PowerPoint, mais du shadowing, de l'observation structurée, de l'analyse d'activité. On ne conçoit pas un composant pour un système qu'on n'a pas observé. Et cela suppose d'architecturer dès l'amont la réversibilité, la redondance et les modalités de retrait. La résilience et la réversibilité ne sont pas des correctifs post-déploiement : Ce sont des variables de conception qui conditionnent le niveau de service de l'algorithme d'IA.

Outil d'évaluation. En complément des métriques classiques (AUC, sensibilité, spécificité), ce cadre permet d'évaluer la viabilité systémique d'un déploiement : compatibilité topologique, bilan temporel net, réduction d'incertitude mesurable, niveau de couplage proportionné, soutenabilité du TCO, existence d'un PRA intégrant explicitement le composant IA. Un modèle avec une AUC de 0,95 et un T_net négatif en RCP n'est pas un bon modèle *in situ*. C'est un bon modèle *in silico*. La distinction est coûteuse et généralement fatale à l'adoption en routine de la solution.

Outil de gouvernance et de dialogue. Entre cliniciens, DSI, RSSI, éditeurs, régulateurs et financeurs, ce cadre fournit un vocabulaire commun. Il permet de dépasser le clivage stérile entre enthousiasme technologique et scepticisme organisationnel. Les tensions observées ne relèvent pas d'une opposition idéologique, mais de dimensions différentes du même système : performance, temps, risque, dépendance, résilience. Les deux camps ont souvent raison. Ils ne parlent simplement pas de la même dimension.

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

Ce que ce cadre n'est pas

Ce n'est pas un outil anti-technologie : C'est un outil anti-déploiement naïf.

Il ne prétend pas remplacer l'évaluation algorithmique qui est un prérequis évident. Il explicite les conditions architecturales de son efficacité réelle. Il n'a pas vocation à être définitif. Il vise à structurer un examen *Ex Ante* des fragilités systémiques et de fournir une structure à la nécessaire évaluation médico-économique de ces solutions IA (et sans IA également).

L'écart entre publication et production n'est pas principalement statistique. Il touche à l'organisation, l'architecture et la résilience. Ignorer ces dimensions condamne une partie significative des projets à échouer non pas par manque de performance, mais par sous-estimation des interdépendances.

La fiche bristol, elle, ne demande pas de cadre d'évaluation en six dimensions. Elle fonctionne. Elle est disponible. Elle ne plante pas. Elle ne peut pas être rançongiellisée. Elle survivra tant que le numérique n'offrira pas un niveau de robustesse équivalent.

Ce cadre est issu d'une réflexion sur l'intégration architecturale de l'IA dans les systèmes de santé, nourrie par le développement de TweenMe et le programme PREDICARE.

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD
AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com