

# Une erreur de prédiction détériore une métrique, une erreur d'intervention détériore un patient

Jumeaux numériques de santé, et la proportionnalité entre revendication, preuve et responsabilité.

## 1. La bonne question

Un jumeau numérique de santé qui affiche une bonne courbe de performance a démontré une chose, et une seule : qu'il prédit. Il n'a démontré ni qu'il explique, ni qu'il simule, ni qu'il peut servir à intervenir sur un patient. Prédire, expliquer, simuler, intervenir sont quatre revendications distinctes, et chacune engage une preuve et une responsabilité qui lui sont propres. La thèse de cet article tient en une phrase : *les revendications d'un modèle doivent être proportionnées à la preuve qu'il fournit et à la responsabilité qu'il engage, et une preuve de prédiction ne couvre ni l'une ni l'autre quand la revendication est d'intervenir.*

Ce n'est pas une critique de l'IA, et il faut le dire avant qu'on ne le lise de travers. Un modèle peut être extraordinairement utile sans expliquer, sans simuler et sans intervenir. Un classifieur qui trie des images de rétine ou priorise une file d'attente rend un service réel sans rien prétendre du mécanisme de la maladie. Le problème n'est jamais qu'un modèle soit « limité ». Le problème est qu'il revendique plus qu'il ne prouve, et qu'en santé cette inflation se paie sur le patient, pas sur une métrique.

Le périmètre doit être posé d'emblée, car « jumeau numérique » recouvre trois familles. Le jumeau *purement statistique* apprend des régularités sans modèle du mécanisme. Le jumeau *mécanistique* repose sur des équations physiologiques explicites, comme en pharmacocinétique fondée sur la physiologie ou en modélisation cardiovasculaire. Le jumeau *hybride* combine les deux. La cible de cet article est la première famille, et la part statistique de la troisième. Un jumeau qui porte ses hypothèses causales dans ses équations a déjà fait une partie du travail qu'on va exiger ; un jumeau qui ne tient que par l'ajustement statistique ne l'a pas fait.

Cet article construit un livrable : une *fiche d'identité du jumeau*, conséquence directe de l'audit de ses dépendances, sans laquelle le système n'est pas gouvernable quelle que

soit sa performance. Le lecteur pressé peut sauter à la section 9 ; il y perdra les raisons, pas la conclusion.

## 2. Quatre revendications : une hiérarchie de responsabilité, pas de connaissance

- Prédire, c'est associer une sortie à une entrée.
- Expliquer, c'est savoir pourquoi cette association tient.
- Simuler, c'est produire des contrefactuels fiables, répondre à « que se passerait-il si l'on changeait le traitement ».
- Intervenir, c'est agir sur un patient réel.

La tentation est de voir là un escalier dont chaque marche supposerait la précédente. Ce serait faux, et un clinicien le réfuterait en une phrase.

La médecine intervient depuis toujours sans expliquer, par démarche empirique. On a prescrit l'aspirine pendant des décennies avant de comprendre son mécanisme, donné du lithium sans savoir pourquoi il stabilise, pratiqué l'anesthésie générale dont le mécanisme reste débattu, employé des anticancéreux historiques dont l'efficacité précédait de loin l'explication. On peut intervenir sans expliquer, expliquer sans pouvoir intervenir, prédire sans comprendre. La relation entre ces verbes n'est pas une hiérarchie de connaissance.

C'est une hiérarchie de responsabilité. En passant de prédire à intervenir, ce qui change n'est pas le degré de compréhension exigé, c'est ce qui supporte l'erreur. Une erreur de prédiction détériore une métrique. Une erreur d'explication compromet un modèle. Une erreur de simulation invalide des hypothèses. Une erreur d'intervention détériore un patient. Le saut fondamental n'est pas cognitif ; il est moral, clinique et juridique. C'est lui qui ordonne les quatre verbes, et lui qui justifie que la preuve exigée s'alourdisse à chaque cran.

Revendication	Question	Preuve minimale	Une erreur est supportée par
Prédire	Que va-t-il arriver ?	Performance : discrimination et calibration	une métrique
Expliquer	Pourquoi ?	Dépendances déclarées, hypothèses explicites	un modèle
Simuler	Que se passerait-il si ?	Structure causale, domaine de validité	des hypothèses

Intervenir	Que dois-je faire ?	Assurance proportionnée au risque	au patient
------------	---------------------	-----------------------------------	------------

Cette lecture par la responsabilité éclaire l'exception apparente. On peut atteindre « intervenir » sans passer par « expliquer » : c'est exactement ce que fait l'essai contrôlé randomisé, qui établit l'effet d'un traitement sans son mécanisme, en remplaçant la compréhension par le contrôle expérimental. Mais il fournit alors une preuve *de niveau intervention*, expérimentale, à la mesure de la responsabilité engagée. La faute n'est jamais de sauter l'explication. La faute est d'intervenir avec une preuve de prédiction, qui ne porte ni le mécanisme ni l'expérience, et donc pas la responsabilité.

L'amalgame industriel consiste à présenter un système qui a rempli la première ligne comme s'il avait rempli les suivantes. Le glissement se loge dans un verbe : le jumeau « modélise » le patient, « anticipe » la réponse, « teste » des scénarios. Chacun promet une simulation que la preuve fournie n'établit pas. La question gouvernable n'est pas « le modèle est-il performant ? » mais « que cette performance autorise-t-elle, et qui en répond ? ».

### 3. Le pont vers la décision : la calibration

Une chose étonne dans la plupart des évaluations de jumeaux : la place démesurée de la discrimination, et la portion congrue laissée à la calibration. C'est l'inverse qu'il faudrait, car pour passer de la prédiction à la décision, la calibration compte plus que la discrimination.

La discrimination mesure une capacité de tri : le modèle place-t-il les patients les plus à risque au-dessus des moins à risque ? C'est ce que résumant l'AUC ou le c-index. La calibration mesure autre chose : les probabilités annoncées correspondent-elles aux fréquences observées ? Quand un modèle prédit un risque de vingt pour cent, l'événement survient-il bien chez un patient sur cinq ? Un modèle peut trier parfaitement et se tromper systématiquement sur les niveaux. AUC élevée ne veut pas dire décision fiable.

Or une décision clinique ne s'appuie pas sur un rang, elle s'appuie sur un seuil de probabilité : traiter au-dessus de tel risque, surveiller en dessous. Un seuil n'a de sens que sur des probabilités calibrées. Un modèle qui discrimine bien mais calibre mal enverra les mauvais patients de l'autre côté du seuil, avec une assurance trompeuse, et c'est souvent ce que le changement de distribution dégrade en premier. Le pont entre la prédiction et l'intervention passe par la calibration ; la discrimination seule ne le franchit pas.

## 4. Le signal et la rareté : des facteurs aggravants, pas la racine

On pourrait croire que la racine du problème est la pauvreté du signal clinique. C'est un facteur aggravant, à situer sans s'y attarder. Trois faits relatifs à la théorie de l'information bornent ce qu'un modèle peut faire :

- Aucune transformation d'une donnée n'augmente l'information qu'elle contient sur une cible,
- Un modèle ne crée pas d'information, il en extrait [Cover & Thomas],
- Et à bruit donné, il existe une borne au-delà de laquelle aucun traitement ne récupère davantage de signal [Neyman-Pearson].

*Créer* de l'information est impossible, *extraire* une structure latente est possible, *estimer cette structure de façon fiable* est conditionnel au signal et au nombre de patients.

Quand le signal est faible, l'estimation devient fragile, et le régime de haute dimension et faible échantillon, fréquent en santé, aggrave les choses : avec beaucoup plus de variables que de patients, des structures illusoire paraissent stables [Hall, Marron & Neeman 2005 ; synthèse clinique à compléter]. Mais ce régime ne *fonde* pas la hiérarchie des verbes ; il la rend plus urgente. La preuve : on peut le retirer sans que la thèse cède. « Mon jumeau repose sur quinze millions de patients, et mon modèle d'imagerie a un signal très fort. » Soit. En électrocardiographie, en radiologie, en anatomopathologie, l'apprentissage profond atteint des performances élevées sur des signaux riches et des cohortes massives. Et le problème demeure identique : une discrimination excellente sur des images ne démontre ni le mécanisme, ni l'effet d'une intervention, ni la validité d'un contrefactuel. Quinze millions de patients établissent une association très fiable ; ils n'établissent ni l'explication, ni la simulation, ni le droit d'intervenir. Le signal fort déplace le système plus haut sur la première ligne du tableau ; il ne le fait pas changer de ligne.

Le réel rappelle ce mur quand un design clinique le prend au sérieux : un essai-plateforme adaptatif comme OCTOPUS/PLATYPUS, dans la sclérose en plaques progressive, n'existe que parce que l'effet à mesurer est petit et la cohorte coûteuse [ms-octopus.mrcctu.ucl.ac.uk]. Un design prudent est une confession métrologique. On ajoutera seulement que, côté données, beaucoup de mesures par patient ne font pas beaucoup de patients indépendants : la quantité qui compte est la taille d'échantillon effective, plus proche du nombre de patients que du nombre de lignes.

## 5. Déclarer les dépendances

Si la performance est l'entrée d'un dossier, la première pièce répond à une question : de quoi cette performance dépend-elle ? On y répond par ablation : on retire une à une les sources susceptibles de la porter, et on mesure ce qu'elle perd.

Quatre sources se distinguent, car « prior » recouvre des choses différentes :

- Les équations injectées (prior mécanistique, comme dans les réseaux informés par la physique, déjà employés en régime de petites données [proceedings.neurips.cc, 2024]),
- Les régularités d'un pré-entraînement massif (prior statistique, comme dans MedGemma [developers.google.com, 2025]),
- La forme de l'espace latent appris (prior géométrique),
- Les nomenclatures (prior ontologique).

L'écart entre la performance brute et la performance après retrait d'une source mesure une quantité réelle, qu'il faut nommer correctement : une *dépendance*, pas une *provenance causale*. Les sources interagissent. Trois priors peuvent produire ensemble une performance qui s'effondre dès qu'on en retire un seul ; alors la somme des contributions dépasse le total, l'attribution dépend de l'ordre des ablations, et plusieurs décompositions coexistent sans qu'aucune ne soit la bonne. C'est un problème de répartition voisin des valeurs de Shapley, avec leurs hypothèses propres et leur coût de calcul [Shapley 1953 ; Lundberg & Lee 2017]. La matrice d'ablation mesure ce dont le modèle *dépend* aujourd'hui, pas ce qui *produit causalement* sa performance. Un lecteur formé à la statistique l'objectera, et il aura raison.

Cette honnêteté coûte un test qu'on aurait voulu décisif. On pourrait croire qu'il suffit, pour réfuter l'idée d'une performance importée, de montrer un gain obtenu sans ajout d'information ni réduction de dimension. Le test est faible : une meilleure architecture produit un gain sans ajouter d'information, en réduisant l'éventail des hypothèses possibles, ce qui est déjà une structure imposée, et passe à travers le test. Il corrobore une dépendance ; il ne prouve pas l'absence de source importée.

L'instrument garde sa valeur. Une forte dépendance à un pré-entraînement n'est pas un défaut, c'est une déclaration à produire. Ce qui reste une faute, c'est l'absence de la déclaration, ou un effondrement hors distribution que personne n'a regardé. Cette déclaration, mise en forme, est la fiche de la section 9 : la fiche n'est pas une idée séparée de l'audit, elle en est la sortie.

## 6. La provenance n'est pas la transférabilité

Il reste à défaire une association tenace : l'idée qu'une performance fortement importée serait, pour cette raison, fragile. C'est faux, et l'exemple le plus visible du domaine le montre. AlphaFold dépend d'un prior exogène gigantesque, et sa robustesse dépasse celle de bien des modèles entraînés sur les seules données d'une tâche étroite. Importer une régularité n'est pas une faiblesse ; en santé, c'est souvent ce qui sauve, quand la cohorte du patient est trop petite pour la fournir elle-même.

D'où une distinction qui tranche : la provenance *explique*, la transférabilité *décide*. Savoir d'où vient la performance aide à comprendre pourquoi un modèle tient ou cède. Savoir s'il tient hors de sa distribution d'entraînement et par sous-population est ce qui autorise, ou non, le déploiement. Les deux dimensions sont découplées : une performance peut être largement importée et très transférable, ou largement « maison » et s'effondrer au premier changement de site. Le critère décisif n'est pas la provenance, c'est la stabilité sous changement de distribution. L'audit de dépendance ne sert pas à purger l'exogène au nom d'une pureté endogène ; il sert à le déclarer et à le tester là où il compte.

## 7. Simuler quoi, dans quel domaine de validité

Un jumeau prétend rarement se contenter de prédire ; il prétend simuler. Et la simulation exige autre chose qu'une bonne performance : une structure causale suffisamment correcte. L'énoncé doit être tenu sans atténuation : *même à signal fort, un jumeau peut être faux s'il simule sur une structure causale incorrecte*.

Reste à dire ce que « suffisamment correcte » veut dire. Un modèle causal structurel décrit les variables et leurs relations, et permet de distinguer ce qu'il peut soutenir : une association, l'effet d'une intervention, ou un contrefactuel, les trois barreaux de l'échelle de Pearl [Pearl, *Causality*].

Deux propriétés conditionnent la portée d'une simulation :

- La *transportabilité* dit si un résultat établi dans une population vaut dans une autre, et sous quelles conditions [Bareinboim & Pearl],
- L'*invariance* dit si une relation se maintient à travers les environnements, l'indice le plus fiable qu'elle soit causale et non accidentelle [Peters, Bühlmann].

Une structure causale suffisamment correcte est une structure dont on a montré qu'elle transporte et reste invariante sur le domaine où l'on veut simuler. Hors de ce domaine, la simulation n'est pas fautive par malchance ; elle est hors garantie.

D'où trois niveaux, pas deux états : la *corrélacion prédictive*, sans prétention causale ; la *causalité interventionnelle locale*, qui prédit l'effet de certaines interventions dans un

domaine étroit ; la *causalité structurelle généralisable*, qui capture un mécanisme transportable.

La plupart des systèmes thérapeutiques sérieux vivent au deuxième niveau, sur un domaine borné : ils n'imitent pas un patient, ils éclairent certaines interventions précises.

La faute consiste à vendre ce deuxième niveau comme le troisième. La question n'est donc pas « simule-t-il bien ? » mais « simule-t-il quoi, à quelle échelle, et dans quel domaine de validité ? ». Ces questions devraient être posées dès la conception du modèle. Toute simulation implique en effet un choix de périmètre, de résolution et de niveau d'abstraction (souvent dicté par la structure en quantité et qualité des données disponibles). Pourtant, un biais analytique fréquent conduit à poursuivre l'idéal d'un jumeau numérique universel capable de reproduire l'ensemble du patient. Or la valeur d'un modèle ne dépend pas de son exhaustivité, mais de l'adéquation entre ce qu'il prétend représenter et la décision qu'il doit éclairer.

Un exemple situé à la frontière de l'état de l'art fixe l'idée. Les modèles de langage protéique de la famille ESM atteignent une précision structurelle remarquable, et la littérature reconnaît qu'on ne sait pas encore trancher s'ils internalisent des principes physiques du repliement ou s'ils capturent principalement des corrélations évolutives. Si cette question demeure ouverte pour l'un des systèmes bio-numériques les plus avancés actuellement disponibles, l'industriel qui affirme que son jumeau « simule le patient » devrait au minimum préciser ce qu'il simule, à quelle échelle, et dans quelles limites cette simulation peut être considérée comme valide.

## 8. La responsabilité commande la preuve

Tout converge vers un principe simple. Ce n'est pas le manque d'information causale qui interdit, à lui seul, d'intervenir sur la foi d'une prédiction. C'est qu'une erreur de prédiction détériore une métrique, là où une erreur d'intervention détériore un patient. La preuve exigible doit donc se régler sur la responsabilité engagée, et non sur la seule performance affichée. C'est un principe de gouvernance avant d'être un principe épistémologique.

L'ingénierie critique l'a traduit en barème. L'aéronautique logicielle gradue ses exigences selon la gravité de la défaillance : la norme DO-178C définit cinq niveaux d'assurance, et le nombre d'objectifs de vérification décroît avec la criticité, de soixante et onze au niveau le plus élevé à vingt-six à un niveau mineur [RTCA DO-178C, 2011]. La preuve est proportionnée à l'impact, donc à la responsabilité. La santé dispose d'un cadre de classification du risque, mais d'aucun barème pour la légitimité de la simulation. Le règlement européen sur l'IA classe en haut risque toute IA intégrée à un dispositif médical marqué CE, soit les classes IIa, IIb et III des dispositifs médicaux et A à D des diagnostics in vitro, avec une échéance fixée à août 2027 pour les classes les plus élevées [Règlement

(UE) 2024/1689 ; (UE) 2017/745 ; (UE) 2017/746]. Il exige robustesse, absence de biais, supervision humaine. Il ne dit rien du niveau causal requis avant d'autoriser une intervention.

C'est le chaînon manquant, et on peut le proposer sans excès : un niveau d'assurance qui croît avec le verbe, parce que la responsabilité croît avec lui. Prédire pour trier demande discrimination, calibration et transférabilité. Expliquer ajoute l'audit de dépendance. Simuler exige en plus un niveau causal établi et un domaine de validité déclaré. Intervenir exige une structure causale dont on a montré qu'elle se transpose, et des conditions explicites de non-déploiement. Le coût de la preuve monte avec le verbe parce que le coût de l'erreur, et le nom de qui la porte, augmentent avec lui.

## 9. La fiche d'identité du jumeau

Voici le livrable, et il n'est rien d'autre que l'audit précédent mis en forme. Plutôt qu'une liste de contrôle qu'on oublie à moitié, quatre blocs qu'on retient :

- *L'objet* : la cible prédictive et la cible décisionnelle, qui ne sont pas toujours la même,
- *Les données* : le signal disponible pour cette cible, la taille d'échantillon effective, et les dépendances déclarées par l'audit,
- *La validité* : la transférabilité hors distribution, le niveau causal atteint, le domaine au-delà duquel le jumeau ne doit pas simuler,
- *L'assurance* : le niveau visé, la responsabilité assumée, et les conditions de non-déploiement.

**Objet, données, validité, assurance : ce qu'on prétend, avec quoi, jusqu'où, et sous quelle garantie.**

Une fiche n'a de valeur que si elle ne ment pas sur sa propre précision. C'est le défaut classique des métrologies de gouvernance : afficher des valeurs plus solides qu'elles ne le sont. Une fiche mature renseigne donc chaque entrée par une valeur, un intervalle d'incertitude et un niveau de confiance. Gouverner une estimation fragile comme un fait est la faute symétrique de l'amalgame des verbes : dans un cas on confond prédire et intervenir, dans l'autre estimer et savoir.

Cette fiche n'est pas une norme en vigueur ; c'est une proposition qui rend opérationnel l'esprit des règlements existants pour le cas des jumeaux. Qualifier un jumeau de « candidat à tel niveau d'assurance » n'est pas le déclarer conforme, et reste un cadre doctrinal, pas un avis juridique.

## 10. Limites

L'audit de dépendance ne livre pas la provenance causale de la performance : en présence d'interactions, l'attribution fine est indécidable sans hypothèses fortes. Le test de réfutation proposé est partiel : une meilleure architecture produit un gain sans information ajoutée, et il corrobore sans réfuter. Les instruments ont un coût que cet article ne chiffre pas : le barème dit que la preuve doit être proportionnée au risque, il ne dit pas où placer exactement les seuils. Le cadre suppose qu'on sache nommer la cible décisionnelle, ce qui reste souvent implicite. La répartition fine de la responsabilité déborde l'épistémologie et relève d'un travail juridique distinct, hors périmètre ici : cet article montre que chaque verbe en appelle un régime différent, il ne l'instruit pas. Enfin, la critique vise le jumeau majoritairement statistique/probabiliste ; un jumeau mécanistique qui porte ses hypothèses causales dans ses équations mérite une évaluation distincte, sur la justesse de ces équations plutôt que sur la déclaration de ses dépendances.

## 11. Conclusion

Le débat dominant demande : le modèle est-il performant ? C'est la première d'une série de questions qu'on a appris à ne pas distinguer.

Performer autorise à prédire. Cela n'autorise pas à expliquer, simuler ou intervenir sur la seule foi de la même preuve. Ce que cet article propose n'est pas une critique de l'IA mais une règle de proportionnalité : à chaque revendication, le niveau causal, la preuve exigible et la responsabilité qui lui correspondent. La règle paraîtra évidente une fois lue. Elle l'est si peu en pratique que le marché vit de l'ignorer.

Cette règle n'est pas propre aux jumeaux, ni même à la santé. La même proportionnalité gouverne toute revendication algorithmique qui franchit la frontière de la prédiction vers l'action : le scoring bancaire, le véhicule autonome, l'agent logiciel qui agit seul, les systèmes d'armes. Le jumeau de santé n'en est qu'un cas, mais c'est le cas aigu, celui où l'erreur, au bout de la chaîne, n'est pas supportée par une métrique mais par un corps. La santé n'invente pas le principe ; elle est l'endroit où il fait le plus mal de l'ignorer.

D'où la phrase qui résume tout le reste, et la seule qu'il faut retenir si l'on n'en retient qu'une : une erreur de prédiction détériore une métrique, une erreur d'intervention détériore un patient. Un jumeau numérique qui prétend intervenir avec une preuve de prédiction n'est pas un patient virtuel. C'est une décision déguisée en mesure, et une responsabilité que personne n'a signée.

## Références

### *Théorie de l'information et détection*

- Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2<sup>e</sup> éd.). Wiley. [inégalité de traitement de données]
- Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A*, 231, 289-337.
- Van Trees, H. L. (2001). *Detection, Estimation, and Modulation Theory, Part I*. Wiley. [filtre adapté]

### *Haute dimension, faible échantillon*

- Hall, P., Marron, J. S., & Neeman, A. (2005). Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society: Series B*, 67(3), 427-444. [référence méthodologique ; une synthèse clinique de synthèse reste à ajouter]

### *Attribution de contribution*

- Shapley, L. S. (1953). A value for n-person games. In *Contributions to the Theory of Games II* (p. 307-317). Princeton University Press.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions (SHAP). *Advances in Neural Information Processing Systems (NeurIPS)* 30.

### *Jumeaux, priors et modèles*

- Med-Real2Sim: Non-Invasive Medical Digital Twins using Physics-Informed Self-Supervised Learning. *NeurIPS* 2024. [https://proceedings.neurips.cc/paper\\_files/paper/2024/file/0b081a44ed0b8c0c4aa6bd886a60bea4-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2024/file/0b081a44ed0b8c0c4aa6bd886a60bea4-Paper-Conference.pdf)
- Physics-informed neural networks for modeling physiological time series for cuffless blood pressure estimation. *npj Digital Medicine* (2023). <https://www.nature.com/articles/s41746-023-00853-4>
- MedGemma model card. *Google Health AI Developer Foundations* (2025). <https://developers.google.com/health-ai-developer-foundations/medgemma/model-card> ; *MedGemma Technical Report*, arXiv:2507.05201 (2025).

- Jumper, J., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583-589. <https://www.nature.com/articles/s41586-021-03819-2>
- Lin, Z., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model (ESM-2 / ESMFold). *Science*, 379(6637), 1123-1130. <https://www.science.org/doi/10.1126/science.ade2574>
- Protein Language Models Encode Evolutionary Grammar but Conflate Topological and Thermodynamic Phases. *bioRxiv* (2026). <https://www.biorxiv.org/content/10.64898/2026.04.07.717117v1.full> [question ouverte : principes physiques vs corrélations évolutives]

### Causalité

- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2<sup>e</sup> éd.). Cambridge University Press.
- Pearl, J., & Mackenzie, D. (2018). *The Book of Why*. Basic Books.
- Bareinboim, E., & Pearl, J. (2016). Causal inference and the data-fusion problem. *PNAS*, 113(27), 7345-7352. <https://www.pnas.org/doi/10.1073/pnas.1510507113> [transportabilité]
- Peters, J., Bühlmann, P., & Meinshausen, N. (2016). Causal inference using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society: Series B*, 78(5), 947-1012. <https://rss.onlinelibrary.wiley.com/doi/10.1111/rssb.12167> [invariance]

### Terrain et cadre réglementaire

- OCTOPUS — Optimal Clinical Trials Platform for Progressive Multiple Sclerosis. MRC Clinical Trials Unit at UCL. Protocole. <https://ms-octopus.mrcctu.ucl.ac.uk>
- RTCA DO-178C (2011). *Software Considerations in Airborne Systems and Equipment Certification*. [niveaux d'assurance / DAL]
- Règlement (UE) 2024/1689 du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'IA).
- Règlement (UE) 2017/745 relatif aux dispositifs médicaux (MDR).
- Règlement (UE) 2017/746 relatif aux dispositifs médicaux de diagnostic in vitro (IVDR).