

## Bonnes pratiques d'apprentissage machine pour le développement des instruments médicaux : Principes directeurs

Octobre 2021

La Food and Drug Administration (FDA) des États-Unis, Santé Canada et la Medicines and Healthcare products Regulatory Agency du Royaume-Uni ont dégagés ensemble 10 principes directeurs qui pourraient servir de base à l'élaboration de bonnes pratiques d'apprentissage machine (BPAM). Ces principes directeurs contribueront à promouvoir des instruments médicaux sans danger, efficaces et de haute qualité faisant appel à l'intelligence artificielle et à l'apprentissage machine (IA/AM).

Les technologies d'intelligence artificielle et d'apprentissage machine sont susceptibles de transformer les soins de santé en raison de leur capacité à tirer avantage des découvertes importantes qui résultent des grandes quantités de données générées quotidiennement pendant la fourniture de soins de santé. Elles emploient des algorithmes logiciels pour apprendre de l'utilisation en situation réelle et, dans certaines circonstances, peuvent utiliser ces renseignements pour améliorer la performance du produit. Mais elles posent également des enjeux particuliers à cause de leur complexité et du fait que leur développement est itératif et axé sur les données.

Ces 10 principes directeurs visent à jeter les bases de l'élaboration de BPAM qui prennent en compte le caractère unique de ces produits. Ils aideront aussi à stimuler la croissance future dans ce domaine qui connaît une évolution rapide.

Les 10 principes directeurs ciblent les domaines sur lesquels l'International Medical Device Regulators Forum (IMDRF), les organisations internationales de normalisation et d'autres organismes de

collaboration pourraient travailler pour faire avancer les travaux sur les BPAM. Les domaines de collaboration sont, entre autres, la recherche, la création d'outils et de ressources pédagogiques, l'harmonisation internationale et les normes consensuelles, ce qui pourrait aider à orienter l'élaboration de politiques et de lignes directrices réglementaires.

Nous envisageons que ces principes directeurs soient utilisés pour :

- Adopter les bonnes pratiques qui ont fait leurs preuves dans d'autres secteurs
- Moduler les pratiques d'autres secteurs afin qu'elles soient applicables à la technologie médicale et au secteur des soins de santé
- Créer de nouvelles pratiques propres aux technologies médicales et au secteur des soins de santé

À l'heure où le domaine des instruments médicaux ayant recours à l'IA/AM évolue, les meilleures façons de faire et les normes consensuelles en ce qui concerne les BPAM doivent évoluer elles aussi. Des associations solides avec nos partenaires internationaux du secteur de la santé publique seront cruciales si nous voulons donner aux parties prenantes les moyens de faire progresser les innovations responsables dans ce secteur. En conséquence, nous espérons que ce premier travail de collaboration pourra guider notre engagement international plus large, notamment avec l'IMDRF.

Nous vous invitons à continuer à nous faire part de vos commentaires via le registre public ([FDA-2019-N-1185](https://www.fda.gov/oc/2019/11/15/fda-2019-n-1185)) sur Regulations.gov, et nous sommes impatients de collaborer avec vous dans le cadre de ces efforts. Le Digital Health Center of Excellence dirige ces travaux pour le compte de la FDA. Communiquez directement avec nous à [Digitalhealth@fda.hhs.gov](mailto:Digitalhealth@fda.hhs.gov), à [software@mhra.gov.uk](mailto:software@mhra.gov.uk) et à [mddpolicy-politiquesdim@hc-sc.gc.ca](mailto:mddpolicy-politiquesdim@hc-sc.gc.ca).

Bonnes pratiques d'apprentissage machine pour le développement des instruments médicaux : Principes directeurs	
L'expertise multidisciplinaire est mise à profit tout au long du cycle de vie du produit	De bonnes pratiques de génie logiciel et de sécurité sont mises en œuvre
Les participants aux essais cliniques et les ensembles de données sont représentatifs de la population de patients ciblée	Les ensembles de données d'entraînement sont indépendants des données de test
Les ensembles de données de référence sélectionnés reposent sur les meilleures méthodes connues	La conception des modèles est adaptée aux données existantes et reflète l'utilisation prévue d'un instrument
L'accent est mis sur la performance de l'équipe humain-IA	Les tests démontrent les performances de l'instrument dans des conditions cliniquement pertinentes
Les utilisateurs reçoivent des informations claires et essentielles	Les modèles déployés font l'objet d'un suivi des performances et les risques de recyclage sont gérés

## Principes directeurs

- 1. L'expertise multidisciplinaire est mise à profit tout au long du cycle de vie du produit :** Une compréhension approfondie de l'intégration prévue d'un modèle dans le déroulement des activités cliniques, ainsi que des avantages souhaités et des risques encourus par les patients, peut contribuer à garantir que les instruments médicaux compatibles avec l'AM sont sans danger et efficaces et répondent à des besoins cliniquement significatifs tout au long de leur cycle de vie.
- 2. De bonnes pratiques de génie logiciel et de sécurité sont mises en œuvre :** La conception du modèle est mise en œuvre en portant une attention particulière aux principes fondamentaux : bonnes pratiques en matière de génie logiciel, assurance de la qualité des données, gestion des données et des pratiques robustes en matière de cybersécurité. Au nombre de ces pratiques figure un processus méthodique de conception et de gestion des risques qui permet de consigner et de communiquer comme il se doit les décisions et les justifications en matière de conception, de mise en œuvre et de gestion des risques, ainsi que de garantir l'authenticité et l'intégrité des données.
- 3. Les participants aux essais cliniques et les ensembles de données sont représentatifs de la population de patients ciblée :** Les protocoles de collecte de données doivent être conçus de manière à ce que les caractéristiques pertinentes de la population de patients visée (par exemple, leur âge, leur sexe, leur race et leur origine ethnique), l'usage et les entrées de mesure soient suffisamment représentées au sein d'un échantillon de taille adéquate dans l'étude clinique et dans les ensembles de données d'entraînement et de test, afin que les résultats puissent être raisonnablement généralisés pour la population ciblée. Ceci est important pour gérer tout biais, promouvoir des performances appropriées et généralisables dans la population de patients visée, évaluer la facilité d'utilisation et mettre en évidence les circonstances dans lesquelles le modèle peut être moins performant.
- 4. Les ensembles de données d'entraînement sont indépendants des données de test :** Les ensembles de données d'entraînement et de test sont choisis et conservés de manière à être adéquatement indépendants les uns des autres. Toutes les sources potentielles de dépendance, comme les facteurs liés au patient, à l'acquisition des données et au site, sont prises en compte et traitées pour garantir l'indépendance.
- 5. Les ensembles de données de référence sélectionnés reposent sur les meilleures méthodes connues :** L'utilisation des meilleures méthodes connues et acceptées pour créer un ensemble de données de référence (c'est-à-dire un étalon de référence) garantit que des données cliniquement pertinentes et bien caractérisées sont recueillies et que les lacunes de la référence sont comprises. S'ils sont accessibles, les ensembles de données de référence acceptés dans le cadre du développement et de l'essai du modèle favorisant et démontrant la robustesse et la généralisabilité de ce dernier dans la population de patients visée sont utilisés.
- 6. La conception des modèles est adaptée aux données existantes et reflète l'utilisation prévue d'un instrument :** La conception des modèles est adaptée aux données existantes et prend en charge l'atténuation active des risques connus, tels que le surapprentissage, la dégradation des performances et les risques pour la sécurité. Les avantages et les risques cliniques liés au produit sont bien compris, utilisés pour dégager des objectifs de performance cliniquement significatifs pour les tests, et confirmer que le produit peut atteindre son utilisation prévue de manière sûre et efficace. Les facteurs à prendre en compte sont notamment les effets des performances globales et locales ainsi que l'incertitude et la variabilité des entrants et des sortants d'un instrument, les populations de patients visées et les conditions d'utilisation clinique.
- 7. L'accent est mis sur la performance de l'équipe humain-IA :** Si le modèle a une composante d'intervention humaine, les considérations relatives aux facteurs humains et à l'interprétabilité des extraits du modèle par l'humain sont abordées en mettant l'accent sur les performances de l'équipe humain-IA, plutôt que sur celles du modèle pris isolément.
- 8. Les tests démontrent les performances de l'instrument dans des conditions cliniquement pertinentes :** Des plans de test statistiquement valables sont élaborés et exécutés pour produire des données sur les performances d'un instrument qui sont cliniquement pertinentes, indépendamment de l'ensemble des données d'entraînement. Les considérations comprennent la population de patients visée, les sous-groupes importants, l'environnement clinique et l'utilisation par l'équipe humain-IA, les entrées de mesure et les facteurs de confusion potentiels.
- 9. Les utilisateurs reçoivent des informations claires et essentielles :** Les utilisateurs ont facilement accès à des informations claires, contextuellement pertinentes et adaptées au public visé (tels que les fournisseurs de soins de santé ou les patients), notamment : utilisation prévue et le mode d'emploi du produit; performances du modèle pour les sous-groupes appropriés; caractéristiques des données utilisées pour entraîner et tester le modèle; entrées acceptables; limitations connues; interprétation de l'interface utilisateur; intégration du modèle dans le déroulement des activités cliniques. Les utilisateurs sont également informés des modifications apportées à l'instrument et des mises à jour découlant de la surveillance des performances réelles, ainsi que des éléments sur lesquels les décisions sont fondées (selon la disponibilité) et d'un moyen de faire part au concepteur de leurs préoccupations relatives au produit.
- 10. Les modèles déployés font l'objet d'un suivi des performances et les risques de recyclage sont gérés :** Les modèles déployés peuvent être surveillés dans des conditions d'utilisation réelle, l'accent étant mis sur le maintien ou l'amélioration de la sécurité et des performances. En outre, lorsque les modèles font l'objet d'un entraînement périodique ou continu après leur déploiement, des contrôles appropriés sont en place pour gérer les risques de surajustement, de biais involontaire ou de dégradation du modèle (par exemple, dérive de l'ensemble de données) qui peuvent influencer sur la sécurité et les performances du modèle lorsqu'il est utilisé par l'équipe humain-IA.