

# L'évaluation des dispositifs médicaux, entre apprentissage automatique et transparence algorithmique.

Audrey DEQUESNES, Doctorante au CERAPS (UMR 8026), Université de Lille

## 1. Communauté d'appartenance

Je suis doctorante en droit public, ma communauté d'appartenance dans l'organisation cet évènement est donc celle des juristes. Je suis plus particulièrement intéressée par les questions du numérique dans la santé, ayant une formation initiale de Pharmacie. Dans le but de mêler les éléments techniques et applicatifs à la recherche en Droit, j'ai suivi une formation universitaire sur l'Intelligence Artificielle en santé, afin d'acquérir quelques bases théoriques d'apprentissage automatique, de classification, de programmation, ou encore de *data mining*.

## 2. Contribution

Avec l'entrée en vigueur du Règlement EU n°2017/745 relatif aux dispositifs médicaux<sup>1</sup>, les organismes notifiés sont chargés de valider la conformité des dispositifs médicaux à travers le marquage CE. Le règlement spécifie que « *les dispositifs comportant des systèmes électroniques programmables, notamment des logiciels, ou les logiciels qui sont des dispositifs à part entière sont conçus de manière à garantir la répétabilité, la fiabilité et les performances eu égard à leur utilisation prévue.* » La méthode pour garantir la répétabilité, la fiabilité et les performances n'est pas explicitée, et si en théorie ces critères peuvent être utilisés pour évaluer des logiciels comme tout autre dispositif, en pratique l'utilisation de l'apprentissage automatique vient altérer la faisabilité de cette évaluation. Le Règlement n'aborde la question de la modification d'un logiciel que dans le cas où interviendrait une modification considérée comme majeure portant sur les performances initiales, la sécurité ou l'utilisation prévue du logiciel, ou l'interprétation des données. « *Ces modifications peuvent consister en algorithmes nouveaux ou modifiés ou peuvent concerner les structures de la base de données, la plateforme d'exploitation, l'architecture, de nouvelles interfaces utilisateurs ou de nouveaux canaux d'interopérabilité* »<sup>2</sup>. Il est attendu du fabricant qu'il demande un nouvel identifiant pour son dispositif médical dans le cas où l'une de ces modifications interviendrait. Il s'agit donc vraisemblablement de modifications souhaitées ou du moins contrôlées par le fabricant, ce qui exclut l'apprentissage automatique de cette disposition.

---

<sup>1</sup> « Règlement (UE) 2017/745 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) n° 178/2002 et le règlement (CE) n° 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2017, p. 745.

<sup>2</sup> Source ?

Il semble d'ailleurs que les critères du marquage CE ne soient pas considérés comme suffisants par la Haute Autorité de Santé (HAS) pour garantir la fiabilité clinique des dispositifs, et notamment par sa Commission nationale d'évaluation des dispositifs médicaux et des technologies de santé (CNEDiMTS). En effet, cette commission intervient dans l'évaluation des dispositifs qui possèdent déjà un marquage CE et pour lesquels le fabricant fait une demande de prise en charge par l'assurance maladie. Cette seconde évaluation n'intervient donc pas pour tous les dispositifs médicaux et n'interfère pas avec leur mise sur le marché, mais atteste de leur pertinence clinique. L'analyse de la CNEDiMTS ne repose plus, comme le marquage CE, sur des critères de sécurité d'utilisation, mais propose une appréciation des dispositifs médicaux basée sur l'évidence clinique, et sur leur place au sein de l'arsenal thérapeutique préexistant. La CNEDiMTS reconnaît que l'apprentissage automatique induit un « *changement de paradigme* »<sup>3</sup> dans l'évaluation des dispositifs médicaux. La Commission estime qu'il ne lui revient pas d'évaluer la pertinence du modèle mathématique soutenant le fonctionnement du dispositif, et souhaite maintenir une évaluation des critères cliniques, de l'intérêt pour le patient et pour la santé publique. Toutefois, elle entend établir de nouveaux critères spécifiques aux algorithmes auto-apprenants, pour que l'évaluation reste réaliste tout au long de l'utilisation du dispositif. Huit critères sont envisagés : la finalité d'usage, l'apprentissage, les données d'entrée d'apprentissage initial ou de réapprentissage, les données d'entrée impliquées dans la décision, la performance, la validation, la résilience du système, l'explicabilité / interprétabilité. Pour chacun, la méthode d'évaluation comme les critères d'acceptabilité et scores de confiance restent à établir, dans les cas où ils pourraient l'être. Une approche différente consiste en l'évaluation et l'approbation des algorithmes en eux-mêmes, afin d'en déterminer la qualité indépendamment des résultats cliniques. C'est l'approche choisie par la FDA aux États-Unis<sup>4</sup>, qui a déjà autorisé les premiers dispositifs médicaux basés sur des algorithmes auto-apprenants. Seulement, les critères d'évaluation ne spécifient pas l'origine et le type des données qui doivent être utilisées. Ainsi, la plupart des algorithmes est approuvée sur la base d'études rétrospectives, les études prospectives montrant des résultats moins favorables<sup>5</sup>. De même, le jeu de données utilisé pour l'évaluation n'est pas nécessairement issu d'une cohorte externe à celle qui a servi à l'apprentissage, ce qui soulève d'évidentes questions d'adaptabilité et de réaction aux biais par des données différentes de celles du matériau d'entraînement<sup>6</sup>. Une large méta-analyse des études portant sur l'efficacité des intelligences artificielles en imagerie médicale mène en particulier au constat que moins d'une étude sur mille est conçue avec une méthodologie suffisamment robuste<sup>7</sup>.

Partant de ces constats, comment pourrait-on traduire les critères de l'évaluation des dispositifs médicaux en exigences techniques pour le fabricant lors de sa demande de

---

<sup>3</sup> « Consultation publique sur le projet de grille d'analyse destinée à être utilisée par la CNEDiMTS pour contribuer à son évaluation de dispositifs médicaux embarquant des systèmes décisionnels s'appuyant sur des procédés d'apprentissage automatique (« Intelligence artificielle ») », publié le novembre 2019, [consulté le 18 février 2020].

<sup>4</sup> US FOOD AND DRUG ADMINISTRATION, « Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) », [consulté le 19 février 2020].

<sup>5</sup> LIU Xiaoxuan, FAES Livia, KALE Aditya U *et al.*, « A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging », *The Lancet Digital Health*, 1, 2019.

<sup>6</sup> COOK Tessa S, « Human versus machine in medicine », *The Lancet Digital Health*, 1, 2019.

<sup>7</sup> LIU Xiaoxuan, FAES Livia, KALE Aditya U *et al.*, *op. cit.*

marquage CE ? Les huit points critiques évoqués par la HAS pourraient-ils se matérialiser en éléments concrets et vérifiables et faire l'objet d'un référentiel de méthodes valides et de cotations pour les fabricants, voire d'une certification à part entière ?