

# L'AVENIR DE LA MÉDECINE PRÉDICTIVE<sup>1</sup>

Par Roman Hossein Khonsari



Réfléchir au futur de l'humanité, c'est prédire : en spéculant, avec une efficacité qui dépend de la clairvoyance de l'auteur, ou en modélisant, avec des outils reposant sur des bases théoriques ou sur des données réelles.

Quelques rares spéculateurs visionnaires ont anticipé des révolutions scientifiques grâce à la puissance de leur esprit<sup>2</sup>, mais la plupart des

- 
- 1 L'objet de cet article étant de formuler une prédiction, sa démarche entre en contradiction avec l'ensemble des principes qui y sont énoncés.
  - 2 Anticipation des principes de la génétique moléculaire : E. Schrödinger, *What is life ? The physical aspect of the living cell*, Cambridge University Press, 1944. Anticipation des principes de l'intelligence artificielle : AM. Turing, « *Computing machinery and intelligence* », *Mind*, 1950, n° 49, pp. 433-460. Quelques principes généraux sur l'origine des divinations tombées juste : N. Silver, *The signal and the noise: why so many predictions fail - but some don't*, Penguin Books, Londres, 2015.

devins se sont trompés<sup>3</sup> et continuent de se tromper<sup>4</sup>. Pour espérer bénéficier d'une portée générale, il faut aborder la prédiction comme une science. Cette « anticipation rationalisée » s'est d'abord développée dans des domaines angoissants, dangereux, ou coûteux de l'argent – par exemple l'analyse financière, les phénomènes climatiques, la résistance des matériaux, ou encore l'origine des maladies graves.

En clair, la prédiction scientifique vise en priorité à éviter les catastrophes et à maximiser les profits. En miroir de ces objectifs pragmatiques, cette branche de la science peut être exploitée autrement : les outils de prédiction rationnelle peuvent aider à comprendre l'évolution de phénomènes naturels ou sociaux, depuis les mouvements cellulaires jusqu'aux processus cognitifs, avec une vision positive. L'accès à la prédiction est cependant inégal selon les branches des sciences de la nature. La physique ou la chimie disposent de bases théoriques et formelles consensuelles qui ont permis depuis plusieurs siècles de construire des outils prédictifs rigoureux et utiles, par exemple en mécanique du solide ou en électronique. La situation est bien différente en médecine.

3 Deux exemples édifiants : Nostradamus, *Les premières centuries ou « Prophéties »* (édition Macé Bonhomme, 1555), Droz, 1996, et Ariane Chemin et Marie-France Etchegoin, *Didier Raoult, une folie française*, Gallimard, 2021.

4 Un « grand patron » de neurochirurgie parisien reconnu pour son implication dans l'innovation chirurgicale avait installé, en 2005, de grands écrans de télévision dans son bureau pour surveiller l'activité dans les blocs opératoires de son service, où il avait fait placer des caméras. Il pouvait même échanger avec (vociférer sur) ses adjoints via un micro qui diffusait sa voix dans des batteries de haut-parleurs. Ces écrans de télévision, les derniers avec tubes cathodiques, étaient très volumineux et occupaient presque toute la place dans son bureau. J'étais allé le voir en tant qu'externe intéressé par la recherche en chirurgie. Très fier de la taille et de la qualité (pour l'époque) des images sur ses écrans, il m'avait expliqué : « L'homme a besoin de grands formats, les petits téléphones avec des écrans minuscules n'ont aucun avenir, personne ne regardera des photos ou des vidéos miniatures. » Le premier téléphone portable avec un appareil photo intégré était sorti en 1999 au Japon (Kyocera Visual Phone VP-210) et le premier iPhone 2G allait être mis sur le marché aux États-Unis deux ans plus tard, en juin 2007.

## La prédiction médicale : entre science, société et politique

Qu'elle soit guidée par des principes positifs ou consacrée au dépistage de désastres à venir, la prédiction est au cœur de l'activité médicale : comment définir un calendrier chirurgical ? Que répondre aux demandes d'un patient sur le pronostic de son cancer ? Combien investir dans une politique de vaccination à grande échelle ? Un problème émerge d'emblée : la médecine repose sur des outils issus principalement de la biologie et, partant, ne dispose pas encore de cadre théorique général. La biologie est en effet une branche des sciences de la nature pour laquelle nous n'avons pas encore déterminé de principes unificateurs, en dehors de la seule théorie de l'évolution<sup>5</sup>. En raison de son lien fort avec la biologie et de l'absence d'outils transversaux, la prédiction en médecine fait appel à des méthodes hétérogènes : corrélations phénotype/génotype, épidémiologie, modélisation biomécanique, raisonnements physiologiques. Et le plus souvent, plusieurs de ces approches sont associées, sans la garantie théorique de leur compatibilité.

Par ailleurs, au-delà des questions de méthode, l'application d'outils de prédiction avec des marges d'erreur à un individu conscient ne répond pas à son questionnement personnel, car il se retrouve, dans la plupart des modèles, lâché dans une distribution gaussienne avec l'espoir d'être du bon côté de la courbe, même avec une probabilité négligeable<sup>6</sup>. Dans ce contexte, la prédiction médicale, du moins dans le domaine des événements négatifs, soulève d'emblée un double problème : absence de méthode consensuelle et ambiguïté de sa pertinence à l'échelle individuelle.

.....

5 Même le lien a priori évident entre vie et organisation cellulaire est mis en cause par la structure des virus : T. Hearn, *Infravies, Le vivant sans frontières*, Seuil, 2019. La théorie de l'évolution est considérée comme la seule théorie unificatrice actuellement connue en biologie : Alain Prochiantz, *Accident : Regard sur la République des sciences*, Odile Jacob, 2024.

6 S.J. Gould, « The median isn't the message », *Virtual Mentor*, n° 15, 2013, pp.77-81.

## La donnée de santé : un lien entre la science et la politique

La base de la médecine prédictive est la modélisation, c'est-à-dire la réduction du monde réel à un système contrôlé dont les lois et les paramètres sont choisis par les architectes du modèle. Tous les modèles ne sont pas déterministes, dans la mesure où certaines lois comportent un degré d'incertitude<sup>7</sup>. Cependant, il n'est pas

abusif de considérer que le modélisateur façonne le monde dans lequel il se positionne pour produire ses prédictions. La démarche de modélisation comprend ainsi d'emblée un élément subjectif, qui n'est pas problématique en soi car les règles choisies par l'opérateur sont énoncées, mais qu'il est important de ne pas négliger lors de l'interprétation des prédictions, surtout dans le domaine de la santé<sup>8</sup>. Plus concrè-

“ En médecine, en l'absence de théorie unificatrice, la construction d'un modèle peut suivre deux chemins : mécanistique ou statistique. ”

7 La Quantification des marges et de l'incertitude (*Quantification of margins and uncertainties*, QMU) est une branche des statistiques d'aide à la décision, qui étudie les seuils et les marges soumis à divers degrés d'incertitude issus de la structure des modèles. La QMU est notamment utilisée dans les industries nucléaires et aérospatiales : M. Pilch et al., « Ideas underlying quantification of margins and uncertainties (QMU) : a white paper », Sandia National Laboratories Report, 2006, SAND2006-5001. Pour une réflexion générale sur l'importance de l'étude des équilibres des systèmes dynamiques en biologie : E. Sanchez-Palencia et al., « Dialectique dans les sciences et systèmes dynamiques », Le temps des cerises, 2022.

8 « Tous les modèles sont faux » (*All models are wrong*) : Box GEP. *Science and statistics*. J-Am Stat Assoc 1976 ; « mais certains sont utiles » (*but some are useful*) : Box GEP. *Robustness in the strategy of scientific model building*. In Launer RL et al. (éd.) *Robustness in statistics*. Academic Press : Cambridge MA, 1979, p.202-203. Dans le même ordre d'idée, « la carte n'est pas la chose cartographiée » (*the map is not the thing mapped*) : Bell ET. *Numerology*. Williams & Wilkins : Baltimore MD. p.138 ; « une carte n'est pas le territoire qu'elle représente, mais, si elle est juste, elle possède une structure similaire à ce territoire, ce qui rend compte de son utilité » (*A map is not the territory it represents, but, if correct, it has a similar structure to the territory, which accounts for its usefulness*) : A. Korzybski « *Science and sanity: an introduction to non-Aristotelian systems and general semantics* », Science Press Printing Co. : Lancaster PA, 1933, pp. 747-761.

tement, en médecine, en l'absence de théorie unificatrice, la construction d'un modèle peut suivre deux chemins : mécanistique ou statistique.

Dans le premier cas, une loi est choisie pour décrire le phénomène à prédire, par exemple à l'aide d'équations aux dérivées partielles rendant compte d'un système dynamique, comme en physique ou en chimie<sup>9</sup>. Ces modèles mécanistiques sont rares en médecine car les lois elles-mêmes sont rares : quelques cas sont connus en physiologie, en biologie du développement ou en épidémiologie, mais ce sont des exceptions<sup>10</sup>. Dans le cas des modèles statistiques, la démarche de prédiction part d'un jeu de données, parfois massif, et y recherche des tendances, par exemple par des méthodes de régression. Le choix du modèle est ici celui du type de traitement des données, et non du mécanisme sous-jacent au processus<sup>11</sup>.

La plupart des modèles prédictifs en médecine étant des modèles statistiques, la donnée de santé apparaît comme la matière première critique de toute démarche dans ce domaine<sup>12</sup>. Construire de grandes bases de données pour alimenter ces modèles prédictifs soulève des questions scientifiques, techniques, mais également politiques, à chaque étape de la chaîne que représente ce processus : collection, contrôle de la qualité, hébergement, mise à disposition et partage. Ce besoin de données lie ainsi une démarche motivée

9 Formulation d'une loi générale rendant compte de l'apparition spontanée de formes géométriques à partir d'un mélange homogène de molécules ayant des propriétés spécifiques d'activation et d'inhibition, et soumis à des perturbations mineures aléatoires, par un mécanisme de réaction & diffusion : Alan Turing, « *The chemical basis of morphogenesis* », *Phil Trans Roy Soc*, London B 1952, n°237, pp. 37-72.

10 Equations de Lotka-Volterra rendant compte de la dynamique d'un système prédateur-proie : Alfred J. Lotka, *Elements of physical biology*, Williams & Wilkins, Baltimore, (États-Unis), 1925. Et aussi, Volterra V., « *Variazioni e fluttuazioni del numero d'individui in specie animali conviventi* », *Mem Acad Lincei*, Roma, 1926, N°2, pp. 31-113.

11 Même si le choix d'une méthode de traitement des données peut impliquer un arbitrage sur les mécanismes sous-jacents, comme dans le cas de la modélisation de la croissance craniofaciale sur la base de données hétérogènes et non-indépendantes.

12 Sur l'importance de la donnée de santé dans la construction de jumeaux numériques : Soudoplatoff S., « *Le jumeau numérique en santé* », *Fondation pour l'innovation politique*, 2023 - <https://www.fondapol.org/etude/le-jumeau-numerique-en-sante/>

par le savoir et la science à des enjeux de société qui sont, soit modérément structurants, comme les réflexions sur la protection de la vie privée et la valorisation, soit bloquants, comme les interrogations sur la souveraineté et la notion de donnée sensible.

### Les jumeaux numériques

Le terme contemporain pour désigner les modèles multi-échelles du corps humain normal et/ou pathologique est « jumeau numérique ». Le jumeau numérique est un objet virtuel qui duplique une fonction, un organe ou même un organisme, et remplace le réel dans une démarche expérimentale<sup>13</sup> ou prédictive. Pour être plus précis, la différence entre un jumeau numérique et un modèle résiderait dans la possibilité constante d'inclure des données dans le modèle et de le modifier, en prenant en compte par exemple l'évolution de l'environnement ou la croissance<sup>14</sup>. Dans cette acception,

les modèles de type jumeau numérique sont particulièrement pertinents pour la pédiatrie. Leur analogie avec les jumeaux au sens biologique n'est pas très heureuse car elle n'a pas de signification plus intéressante que celle d'une notion de « copie conforme ».

**“ Un jumeau numérique est plus proche, sur le plan conceptuel, d'un schéma, d'une ombre ou d'une silhouette. ”**

En effet, la formation des jumeaux biologiques n'implique en rien une étape de réduction

correspondant à la démarche de modélisation, au cœur de la création d'un jumeau numérique. Par ailleurs, la distinction entre « vrais »

13 Nombreux exemples d'applications en cardiologie et en oncologie, avec notamment des références d'expériences pharmacologiques in silico : Laubenbacher R et al., « Digital twins in medicine », *Nat Comput Sci*, 2024, n°4, pp. 184-191.

14 Wright L et al., "How to tell the difference between a model and a digital twin", *Adv Model Simul Eng Sci*, 2020, n°7, p. 13.

(homozygotes) et « faux » (hétérozygotes) jumeaux n'a pas d'équivalent dans le domaine de la modélisation. Un jumeau numérique est ainsi plus proche, sur le plan conceptuel, d'un schéma, d'une ombre ou d'une silhouette – en clair, d'un objet issu d'une démarche impliquant une force extérieure qui sélectionne des paramètres d'intérêt sur la base de critères subjectifs (schéma) ou physiques (ombre, silhouette), tout en causant de prévisibles mais d'inévitables déformations<sup>15</sup>.

Le domaine des jumeaux numériques est porteur en médecine prédictive. Sur le plan institutionnel, des efforts importants ont été déployés pour le promouvoir, avec des consortiums copieusement financés impliquant des acteurs publics et privés, témoignant de la prise de conscience à de hauts niveaux de responsabilité du caractère stratégique de ces objets digitaux<sup>16</sup>. La notion de jumeau numérique ne présage pas du choix d'un arsenal théorique spécifique et n'est pas plus précise que la notion générale de modèle ouvert appliquée au corps humain. Deux cas d'usage vont permettre de délimiter ce que peut représenter un tel objet.

### **Les jumeaux numériques craniofaciaux : cas d'usage**

La chirurgie craniofaciale est la branche de la médecine dédiée à la correction des malformations touchant à la fois la boîte crânienne et le visage, comme les faciocraniosténoses ou les fentes faciales. Les pionniers de cette discipline, comme Paul Tessier, ont développé des techniques chirurgicales audacieuses à une période (les années 1970) où la pratique des soins comprenait une part inhérente d'expérimentation<sup>17</sup>. La chirurgie a aujourd'hui atteint un très haut

.....  
15 Korzybski, *op. cit.*

16 Exemple de jumeau numérique auquel contribue une entreprise française (Dassault Systèmes) au premier plan international dans ce domaine : M. Peirlinck et al., « Precision medicine in human heart modeling: perspectives, challenges, and opportunities », *Biomech Model Mechanobiol*, n° 20, 2021, pp. 803-831.

17 D. Marchac et al., « Midface surgery from Tessier to distraction », *Childs Nerv Syst*, 1999, n°15, pp. 681-94.

niveau d'efficacité et les marges de progression sont de plus en plus faibles, ne permettant pas d'imposer aux patients des choix intuitifs potentiellement désastreux. Dans ce contexte, le développement de jumeaux numériques est important pour l'évolution de la qualité des soins.

La chirurgie craniofaciale est surtout pédiatrique : la démarche de modélisation doit prendre en compte les modifications de forme et de fonction au moment de l'intervention, mais également estimer la croissance post-opératoire. La croissance craniofaciale est un domaine où les théories sont nombreuses, souvent contradictoires, et les données plutôt rares<sup>18</sup>. Parmi toutes les théories inclusives de la croissance craniofaciale, l'hypothèse des matrices fonctionnelles forme l'une des bases les plus intéressantes pour la construction d'un jumeau numérique.

Cette hypothèse, formulée par Melvin Moss à la fin des années 1960<sup>19</sup>, propose un lien fort entre la formation du squelette craniofacial et les tissus mous qui entourent les pièces squelettiques : ainsi l'expansion massive du cerveau durant la première année de vie modèle la forme de la voûte crânienne, l'augmentation de volume des globes oculaires permet la croissance des orbites, et la croissance de la langue guide l'allongement de la mandibule. L'hypothèse de Moss est séduisante car chacun de ces modules peut bénéficier d'une approche de modélisation de type jumeau numérique, avec la prise en compte d'un jeu de données spécifiques : dynamique de l'augmentation de volume du cerveau, courbes d'expansion des globes oculaires, relevé de la croissance de la langue et de sa projection.

Dans un second temps, ces modules peuvent être intégrés dans un jumeau numérique général de la tête et du cou avec la prise

18 J. Bouaoud et al., « Croissance craniofaciale », *EMC – Odontologie*, 2019, 28-020-H-05.

19 ML. Moss, « *The functional matrix* », in B. Kraus et al., *Vistas in orthodontics*, Lea & Febiger, Philadelphia, 1962, pp. 85-98.

en compte de leurs interactions, liées aux complémentarités des fonctions et à la proximité anatomique des structures dans le massif craniofacial<sup>20</sup>. En effet, ces modules fonctionnels présentent des zones de contact, qui leur imposent des interactions lors de leur croissance spécifique. Ces imbrications anatomiques et fonctionnelles sont des défis de modélisation stimulants et justifient une approche en deux temps : modulaire, puis intégrée.

“ La construction des modèles numériques, déjà riche en défis techniques, dépend ainsi de la disponibilité de données de grand volume et de bonne qualité. ”

Une étape supplémentaire dans la construction du jumeau numérique serait d'attribuer des propriétés physiques à ces modules, d'ordre biomécanique, afin de pouvoir simuler les effets de la chirurgie. Le module fonctionnel de la voûte crânienne a presque atteint, d'après les dernières publications dans ce domaine, un tel niveau d'avancement, combinant géométrie et mécanique : sa « découpe » selon diverses lignes d'ostéotomie permet de prédire les effets morphologiques de diverses approches techniques dans le traitement des malformations craniofaciales.

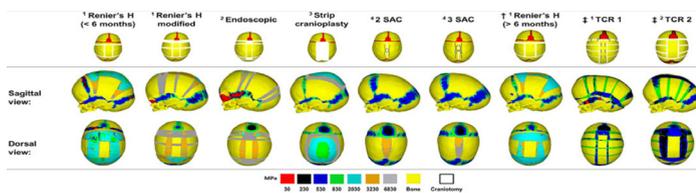
Par exemple, la Figure n°1, ci-dessous, présente 9 découpes testées avec croissance post-opératoire *in silico* durant 5 ans et comparaison avec les données de vie réelle, qui confirment une bonne correspondance morphologique.

À noter que la pertinence clinique réelle de ces jumeaux sera atteinte quand les résultats morphologiques seront associés à des données fonctionnelles (bilan neuropsychologique dans le cas de la chirurgie craniofaciale). La construction des modèles numériques, déjà riche en défis techniques, dépend ainsi de la disponibilité de

20 La prise en compte des interactions entre matrices fonctionnelles a été formalisée sous le nom de principe des contreparties d'Enlow (*Enlow's counterpart principle*) : DH. Enlow *et al.*, *Essentials of facial growth*, WB Saunders Company, Philadelphia PA, 1996, 318 p.

données de grand volume et de bonne qualité. À noter ici que les données disponibles pour la construction des jumeaux numériques craniofaciaux sont exclusivement des données de vie réelle rétrospectives et hétérogènes. Lors de la conception de jumeaux numériques, la situation la plus fréquente est un jeu hétérogène de données non indépendantes, c'est-à-dire que plusieurs échantillons sont disponibles pour plusieurs individus d'une même cohorte, prélevés à des intervalles non contrôlés.

Ce type de données nécessite un arbitrage sur les méthodes d'analyse avec des choix qui ne sont pas anodins en termes d'hypothèses mécanistiques (voir note 11) : ainsi, une hypothèse possible dans les maladies homogènes sur le plan génétique est de considérer que l'échantillonnage de plusieurs scanners à différents âges chez différents individus rend compte du parcours de croissance standard d'un individu type caractéristique de la pathologie<sup>21</sup>. Le recours à cette hypothèse très forte met l'accent sur les effets structurants du choix des méthodes d'analyse, même quand les modèles sont essentiellement statistiques.



<sup>1</sup> Adapted from the Department of Maxillofacial Surgery and Plastic Surgery, Nether - Enfants Malades University Hospital Assistance Publique - Hôpitaux de Paris School of Medicine (Paris, France)  
<sup>2</sup> Adapted from the Department of Head and Neck Surgery for Children and Adolescents, University of Wisconsin and Markey (Chgoyn, France)  
<sup>3</sup> Adapted from the Cranial Craniofacial Unit, Oxford University Hospital (Oxford, UK)  
<sup>4</sup> Adapted from Department of Plastic Surgery, Sahlgrenska University Hospital (Gothenburg, Sweden)

1: Intervention at 6 months  
 2: Intervention at 12 months

*Jumeau numérique craniofacial prédisant l'évolution de la forme du crâne sur une période de 6 ans après neuf types de cranioplasties, avec différents tracés ou pratiqués à différents âges, pour la correction de la scaphocéphalie par fusion prématurée de la suture sagittale (l'une des malformations craniofaciales les plus fréquentes)<sup>22</sup>.*

21 Exemple d'utilisation d'un modèle mixte pour prédire les effets de la chimiothérapie et de la radiothérapie sur le volume de tumeurs cérébrales à partir d'une cohorte de patients disposant de données non indépendantes et hétérogènes : Ribba et al., « A tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy », *Clin Cancer Res*, 2012, n°18, pp. 5071-80. L'incorporation de données d'imagerie dans de tels modèles mixtes reste encore un défi théorique.

22 D'après Cross et al., « A computational framework to predict calvarial growth : optimizing management of sagittal craniosynostoses », *Front Bioeng Biotechnol*, 2022, n° 10, 913190.

## Jumeaux numériques craniofaciaux et intelligence artificielle

Le terme d'intelligence artificielle désigne un éventail large et peu circonscrit de méthodes d'analyse de données qui consistent à reproduire certaines des capacités cognitives animales (classification ou aide à la décision par exemple)<sup>23</sup>. Parmi ces méthodes, les approches par apprentissage profond (*deep learning*) sont très

troublantes car elles ont des propriétés génératives difficiles à anticiper et décrypter, mimant le fonctionnement du système nerveux central humain dans ses tâches considérées comme les plus élevées (création artistique par exemple)<sup>24</sup>.

“ Les algorithmes d'apprentissage profond peuvent détecter des informations qui échappent à nos moyens humains de détection : des associations que nous ne percevons pas, ou des signaux faibles. ”

Le nom de « réseau de neurones » qui leur est associé se réfère à la volonté de leurs concepteurs d'imiter l'organisation des connexions neuronales du cerveau dans la manière dont circule l'information dans l'algorithme<sup>25</sup>. Cette structure permet des modifications du « réseau » en fonction de ce qu'il reçoit comme informations et renvoie à la dé-

finition de jumeau numérique, avec la possibilité de transformer le modèle selon la nature du flux de données entrantes<sup>26</sup>. Cependant, les algorithmes d'apprentissage profond constituent des jumeaux

23 Sur la reproduction algorithmique du fonctionnement neuronal appliquée à la vision : J. Petitot, *Neurogéométrie de la vision : modèles mathématiques et physiques des architectures fonctionnelles*, Les éditions de l'École polytechnique, Palaiseau, 2008.

24 Sur la notion de créativité – synthèse de données massives ou fulgurance ? : M. du Sautoy, *The creativity code, art and innovation at the age of AI*, Belknap Press, Cambridge MA, 2020.

25 Revue sur les principes de l'apprentissage profond : A. Fourcade et al., « Deep learning in medical image analysis: a third eye for doctors », *J Stomatol Oral Maxillofac Surg*, 2019, n° 120, pp. 279-288.

26 Wright et al., *op. cit.*

numériques « parallèles », non humains, dans le sens où ils remplissent sur demande des tâches en imitant en apparence notre fonctionnement cognitif, mais selon des logiques qui, si elles sont fondées sur des données d'entrée réelles, ne correspondent pas obligatoirement à des mécanismes neuropsychologiques naturels. C'est pour cette raison que les algorithmes d'apprentissage profond peuvent détecter des informations qui échappent à nos moyens humains de détection : des associations que nous ne percevons pas, ou des signaux faibles<sup>27</sup>.

Par ailleurs, pour des raisons de technique informatique, dans le domaine de l'apprentissage profond, les étapes allant de la donnée d'entrée à la décision ne sont pas faciles à décrypter (on parle d'un effet « boîte noire »). L'exploitation pratique de phénomènes naturels dont le mécanisme est mystérieux n'a jamais été un frein pour l'homme : l'architecture et l'astronomie se sont développées avant la formulation des lois de l'attraction universelle, et l'utilisation du courant électrique a précédé la compréhension des lois de l'électrodynamique et de la mécanique quantique. Le cas de l'apprentissage profond est toutefois particulier car il s'agit d'un objet entièrement construit par l'homme dont le fonctionnement comporte une part d'inconnu : sur ce point, il représente une rupture conceptuelle avec toutes les autres approches précédentes de modélisation.

L'application de l'apprentissage profond à l'étude des malformations craniofaciales permet entre autres questions d'aborder le problème du diagnostic des syndromes rares. L'errance diagnostique est une source de mortalité et de morbidité importante pour les patients atteints de maladies rares. En France, trois millions de personnes souffrent d'une des 7 000 maladies rares identifiées. Une personne atteinte sur quatre attend plus de quatre ans le début de la recherche de son diagnostic. Une fois dans le circuit du parcours diagnostique, cette recherche dure au moins un an et demi pour

27 Sur la notion de troisième œil offert aux médecins par les algorithmes d'apprentissage profond : Fourcade et al., *J Stomatol Oral Maxillofac Surg*, 2019, n° 120, pp. 279-288.

une grande majorité de malades et dépasse cinq ans pour plus d'un quart d'entre eux. Or, 50% des patients sont des enfants de moins de cinq ans, avec des enjeux majeurs de diagnostic précoce pour leur assurer un développement normal, sans séquelles.

Ces maladies rares sont par ailleurs responsables de 10% des décès survenant entre un et cinq ans. La majorité (80%) des maladies rares sont d'origine génétique, et 30 à 40% des maladies génétiques rares présentent une anomalie craniofaciale<sup>28</sup>, ce qui représente plus de 2 000 maladies différentes, parfois difficiles à identifier, même pour les spécialistes les plus expérimentés<sup>29</sup>.

La compétence en dysmorphologie est rare et repose sur un savoir individuel difficile à partager, résultant de l'analyse de milliers de cas cliniques. Comme les maladies concernées sont rares, les dysmorphologistes compétents sont concentrés là où les malades convergent, dans les centres les plus spécialisés. Les algorithmes d'intelligence artificielle entraînés avec des données photographiques et cliniques sont une solution adaptée pour améliorer les performances diagnostiques de tous les centres prenant en charge des maladies rares. Le succès d'une telle approche dépend de la quantité et de la qualité des données d'entrée.

Ainsi, le projet AIDY (*Artificial Intelligence for Dysmorphology*) entre l'Institut Imagine et l'Assistance Publique – Hôpitaux de Paris, consiste à exploiter une base historique de près d'un million de photographies d'enfants présentant des maladies rares, avec pour la plupart un diagnostic d'expert ou les résultats d'investigations génétiques, afin de développer des outils d'aide à la décision<sup>30</sup>.

28 Pour apprécier la diversité des malformations craniofaciales : R. Hennekam et al., *Gorlin's syndromes of the head and neck* (5<sup>e</sup> édition), Oxford University Press, 2010.

29 ERRADIAG, l'errance diagnostique dans les maladies rares. Enquête de l'Alliance Maladies Rares, 2016. Document disponible sur : <https://www.alliance-maladies-rares.org>

30 Exemple d'utilisation d'AIDY pour l'étude d'une maladie rare craniofaciale, le syndrome Kabuki : Hennocq et al., « *Next generation phenotyping for diagnosis and phenotype-genotype correlations in Kabuki syndrome* », *Sci Rep*, 2024, n°14:2330.

AIDY est un exemple concret du jumeau numérique d'un dysmorphologue, libre de développer une expérience internationale via des projets multi-centriques, toujours disponible et facile à diffuser.

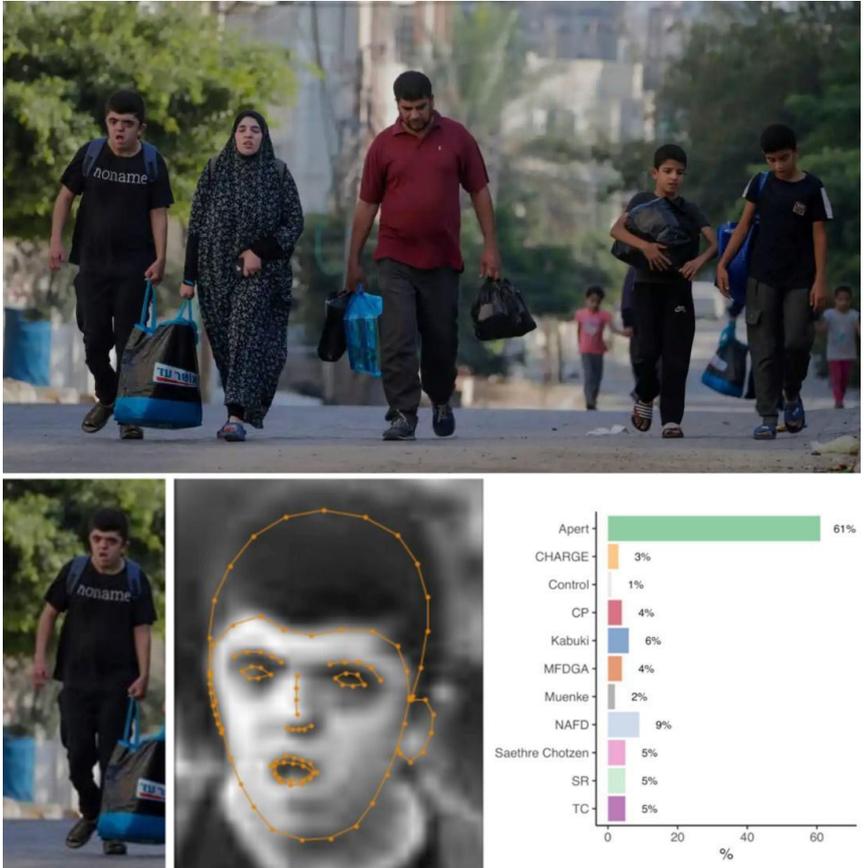
Dans ses utilisations les plus inattendues, AIDY pourrait servir à passer automatiquement en revue les images diffusées librement sur les réseaux sociaux et la presse en ligne, afin de diagnostiquer des pathologies passées inaperçues, dans le domaine des maladies rares et plus largement pour toutes les maladies avec une répercussion sur la forme du visage (comme de nombreuses atteintes endocriniennes ou cardiovasculaires par exemple).

Ce principe d'alerte automatique à grande échelle (Figure 2) s'inspire du journalisme de données (*data journalism*) et reprend des méthodes très répandues dans des sciences où l'observation continue de vastes territoires est cruciale, comme l'astronomie ou l'éthologie, et où des amateurs formés jouent un rôle actif<sup>31</sup>. Une autre utilisation prometteuse d'AIDY est la génération d'images synthétiques<sup>32</sup>, avec la possibilité pour chaque syndrome rare de contrôler le sexe, l'âge, l'ethnie et, dans l'idéal, les antécédents médico-chirurgicaux. Les images synthétiques sont précieuses pour l'enseignement, mais également pour résoudre des problèmes diagnostiques dans des situations inhabituelles liées à l'âge ou à l'origine des patients.

La structuration d'un projet comme AIDY ramène cependant à des questions de société récurrentes liées au développement d'outils utilisant l'intelligence artificielle, et plus généralement à la modélisation en médecine : la protection de la vie privée, la disponibilité des bases de données et leur partage.

31 Sur le rôle des amateurs dans la collecte des données pour l'astronomie et les sciences de l'environnement : Freeman Dyson, *Imagined Worlds*, Harvard University Press, Cambridge MA, 1998.

32 Exemple de générateur de visages synthétiques de la population générale : <https://thispersondoesnotexist.com>



Identification d'un adolescent présentant un syndrome d'Apert sur une photographie diffusée par un quotidien en ligne (*The Guardian*, 15 octobre 2023), dans un reportage relatant les mouvements de population du nord vers le sud de la bande de Gaza. CP = syndrome de Crouzon-Pfeiffer ; MFDGA = dysostose mandibulofaciale type Guion-Almeida ; NAFD = dysostose acrofaciale type Nager ; SR = syndrome de Russel-Silver ; TC = syndrome de Treacher Collins. Photographie originale : Saber Nureldine, European Pressphoto Agency.<sup>33</sup>

<sup>33</sup> D'après Hennocq et al., « Humanitarian facial recognition for rare craniofacial malformations », *Plast Reconstr Surg Glob Open*, 2024 ; 12-e5780.

## Les politiques de la donnée

La matière première de la médecine prédictive est la donnée de santé. Si certains rares praticiens mettent encore en doute la pertinence d'une médecine fondée sur la preuve<sup>34</sup>, la donnée de santé est unanimement considérée comme indispensable à la conduite de soins adaptés et au progrès. L'accès à cette donnée est encore hétérogène sur le territoire, avec des bases dispersées, soumises à des gouvernances variables, à l'interopérabilité inconstante et ne disposant pas toujours d'un contrôle qualité adapté aux questions posées. En termes de définitions, l'utilisation des données de santé pour alimenter des algorithmes de prédiction, ou construire des modèles, rentre dans le domaine de l'utilisation secondaire – l'utilisation primaire désignant l'accès aux données pour les soins apportés aux patients (consultation du dossier médical informatisé par exemple). Par ailleurs, ces données sont en grande majorité des données de vie réelle, ce qui soulève des problèmes spécifiques de qualité et de choix de méthodes d'analyse.

## Où se trouve la donnée ?

La France dispose de plusieurs sources de données structurées pour leur utilisation secondaire. L'une de nos bases les plus emblématiques est le Système national des données de santé (SNDS), dont la base principale (BP) est médico-administrative et gérée par la Caisse natio-

34 Ainsi, dans le rapport soumis en 2020 par Didier Raoult à la Commission d'enquête du Sénat sur l'évaluation des politiques publiques face aux pandémies à la lumière de la crise sanitaire de la Covid-19 et de sa gestion (<https://www.mediterranee-infection.com>), peut-on lire :

à propos de la randomisation : « Les méthodes ne sont que des théories scientifiques comme les autres » ;

à propos des études menées à partir des entrepôts de données : « [...] dans les éléments observationnels, il faut différencier les études observationnelles réalisées par les praticiens, et celles directement prélevées sur des établissements de santé et leurs données informatiques. Dans ce dernier cas, les réponses sont binaires, et, donc, ne correspondent pas à la situation médicale. Par exemple, vous ne pouvez pas traiter des données comme : le patient a-t-il pris de l'Hydroxychloroquine : oui ou non ? En effet, la dose prescrite, la durée de la prescription, le suivi de la prescription, le moment où la prescription a été réalisée, sont des éléments très importants. »

La seconde citation montre une méconnaissance majeure des possibilités offertes par la réutilisation des données de santé en matière de médecine prédictive, notamment par l'exploitation du SNDS et son chaînage avec les EDS.

nale d'assurance maladie (CNAM). La BP du SNDS contient aujourd'hui des données de l'Assurance maladie (base SNIIRAM, qui enregistre par exemple la consommation des médicaments remboursés), des données hospitalières (bases PMSI, qui répertorient par exemples les actes codés), les causes médicales de décès (base CépiDC de l'Inserm). Cette base est très riche, car la quasi-totalité des Français bénéficie de la protection de la CNAM. Elle est en revanche limitée par nature, car elle ne contient pas toujours d'informations directes sur le diagnostic, par exemple quand le patient consulte son médecin traitant en ville. Pour constituer une cohorte d'une pathologie donnée à partir du SNDS, il faut par conséquent développer des algorithmes dédiés qui opèrent des recoupements entre données dans la base pour établir des profils de patients présentant très probablement la pathologie d'intérêt.

L'accès aux données du SNDS n'est autorisé que pour des finalités d'intérêt public (recherche, innovation par exemple). Il est d'abord soumis à l'approbation du Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES), puis à l'autorisation de la Commission nationale de l'informatique et des libertés (CNIL). Certains organismes comme l'Inserm et tous les Centres hospitaliers universitaires (CHU) possèdent cependant un accès permanent au SNDS, qui permet d'extraire des données agrégées et anonymisées et des données personnelles pseudonymisées, dans un périmètre restreint, à la demande des praticiens et des chercheurs, sans autorisation préalable du CNIL et du CESREES.

D'autres sources de données d'intérêt majeur ont une structuration moins avancée que le SNDS à l'échelle du territoire. Parmi elles, les entrepôts de données hospitaliers<sup>35</sup> sont actuellement gérés par les établissements de santé eux-mêmes, parfois regroupés en consortiums, et contiennent des données complémentaires à celles du SNDS (par exemple, résultats des examens radiologiques, données biolo-

35 La Haute Autorité de Santé définit les Entrepôts de Données de Santé comme « la mise en commun des données d'un ou plusieurs systèmes d'information médicaux, sous un format homogène pour des réutilisations à des fins de pilotage, de recherche ou dans le cadre des soins » (<https://www.has-sante.fr>)

giques). Des initiatives à grande échelle avec des financements publics importants tentent aujourd'hui d'uniformiser leur gouvernance, assurer leur interopérabilité et sécuriser leurs plans de financement<sup>36</sup>. Les cohortes et les registres sont également des sources importantes de

données, avec une hétérogénéité majeure dans leurs dimensions et les doctrines déterminant leurs modalités d'accès.

Enfin, des bases liées à des domaines spécifiques de soins sont disponibles, ou en cours de constitution, pour utilisation secondaire, comme Unibase<sup>37</sup>, initiative d'un regroupement de Centres de lutte contre le cancer (CLCC) impliquant 11 sur ces 18 institutions utilisant l'outil ConSoRe d'exploration de données<sup>38</sup>.

Très incomplet, ce panorama

permet toutefois de cerner trois questions soulevées d'emblée par les besoins de réutilisation des données de santé : leur dispersion territoriale et institutionnelle, la diversité de leurs gouvernances, notamment en termes de modalités d'accès, et leur absence d'interopérabilité à grande échelle.

“ Pour optimiser l'utilisation secondaire des données de santé pour la recherche et l'innovation, la France a fait le choix de créer une structure publique dédiée, le *Health Data Hub* ”

36 Dans le cadre de France 2030 et de la stratégie d'accélération « Santé numérique » (SA-SN) lancée en juillet 2022, appel d'offre émis par la Bpifrance : « Accompagnement et soutien à la constitution d'entrepôts de données de santé hospitaliers », suivi par un appel d'offre focalisé sur des projets mobilisant des entrepôts émis en 2023 par la Direction générale de l'offre de soins du Ministère de la santé et de la prévention : « Données de santé et Applications (DATAE) ».

37 Les objectifs d'UNIBASE sont : accélérer l'accès aux données de vie réelle à un grand nombre d'utilisateurs, agréger des données hétérogènes et les intégrer dans un modèle commun, implémenter une solution technique interopérable dans chaque centre, partager les données standardisées et de qualité & mutualiser les compétences (par exemple développer un outil unique d'extraction automatique des images ou unifier l'appariement au SNDS).

38 Pour une présentation de ConSoRe : <https://www.youtube.com/watch?v=leBXmE5Hdzc&t=4s/>

## Centraliser ou fédérer ? La création du Health Data Hub<sup>39</sup>

Pour optimiser l'utilisation secondaire des données de santé pour la recherche et l'innovation, la France a fait le choix de créer une structure publique dédiée, le *Health Data Hub* (HDH, ou Plateforme des Données de Santé), sous la forme d'un groupement d'intérêt public (GIP). À la suite du rapport de Cédric Villani sur l'intelligence artificielle<sup>40</sup>, la loi actant la création du HDH a été promulguée en 2019, poursuivant les réflexions entamées en 2016 avec la création du SNDS par la loi de modernisation du système de santé<sup>41</sup>.

Le HDH est structuré autour de trois notions :

- ▶ un guichet unique clairement identifié par les porteurs de projets, qui permet une concentration des compétences techniques et réglementaires et facilite les démarches quand les données sont dispersées ;
- ▶ un espace de travail de pointe sur le plan technique et sécuritaire, sous forme de nuage (*cloud*), construit en partenariat avec le service de données de santé de Microsoft (Azure) ;
- ▶ un catalogue<sup>42</sup> regroupant des bases d'intérêt majeur, pertinentes sur le plan scientifique et médical, ayant bénéficié d'un important travail de structuration, notamment en termes d'interopérabilité.

L'accès aux données individuelles de santé, et, plus généralement, la conduite de projets dans l'espace de travail opéré par le HDH,

39 Déclaration de conflit d'intérêt : l'auteur est le directeur médical du Health Data Hub, rémunéré à ce titre par ce GIP et détaché de l'AP-HP une journée par semaine dans le cadre d'une mission de service public.

40 Cédric Villani, « Donner un sens à l'intelligence artificielle (IA) », 2018, disponible sur : <https://www.aiforhumanity.fr>

41 Pour un aperçu de l'historique de la création du HDH et de ses bases institutionnelles : <https://www.health-data-hub.fr>

42 Le catalogue du SNDS comprend, entre autres bases, une copie du SNDS dit « historique » (aussi appelé « base principale ») – la CNAM gardant la cogestion de cette base, la Banque Nationale des Maladies Rares (BNDMR), ou encore diverses bases liées au COVID ou à d'autres pathologies : <https://www.health-data-hub.fr/catalogue-de-donnees/>

est soumise au double filtrage de la CNIL et du CESREES. Ce filtrage est justifié par un examen attentif de la méthodologie et des aspects éthique et sécuritaire des projets, mais induit des délais d'accès aux données qui ne sont pas compatibles avec la durée d'une thèse de sciences ou avec les contraintes de survie d'une petite entreprise en recherche d'investissements (*start-up*). La Commission de l'intelligence artificielle a ainsi récemment proposé de réformer le mandat de la CNIL pour l'orienter vers l'innovation et introduire la notion d'accès *a priori* avec sanctions *a posteriori* en cas de mésusage pour les projets d'intérêt public<sup>43</sup>. Ces discussions soulignent l'intrication des enjeux scientifiques, techniques et politiques (surtout dans son sens négatif de jeu de pouvoirs) dès que des décisions s'imposent sur la gestion des données de santé. Cette dernière dimension est cependant centrale pour la réussite des projets de réutilisation des données – et partant, pour le progrès de la médecine prédictive – comme en témoignent les deux polémiques dans lesquelles a été pris le HDH depuis sa création.

La première de ces polémiques concerne la structure même du HDH : le choix de créer un organisme centralisateur, régalién, disposant également d'un rôle opérationnel en termes d'hébergement de données a été vécu comme empiétant sur leur périmètre par de nombreuses structures existantes, comme souvent lors du lancement d'initiatives transversales. Pour beaucoup d'acteurs de la donnée de santé en France, une alternative à cette approche centralisatrice aurait été l'approche fédérée, avec la création de procédures d'accès normalisées pour chaque site hébergeur et la mise à disposition par le HDH d'un catalogue de métadonnées<sup>44</sup>. Ce HDH fédéré aurait alors exercé un rôle administratif pur en émettant des recommandations sur les modalités d'interopérabilité, ou en intervenant partiellement, aux côtés des ministères de

43 Philippe Aghion et al., *IA : Notre ambition pour la France - Travaux de la Commission de l'intelligence artificielle*, Odile Jacob, Paris, 2024.

44 C'est-à-dire d'un catalogue de données documentant les bases de données disponibles dans chaque centre du réseau fédéré.

tutelle, sur la distribution de l'argent public investi dans le soutien à la collecte et l'hébergement des données de santé.

Dans les faits, le catalogue du SNDS dans sa forme actuelle est un choix intermédiaire et raisonnable entre les approches fédérées et centralisatrices, car seule une partie très réduite, mais pertinente, de l'ensemble des données du territoire est destinée à migrer vers le nuage central, et des métadonnées à grande échelle (décrivant les EDS, registres, cohortes entre autres) seront disponibles à terme pour les utilisateurs. L'approche fédérée stricte aurait très probablement nécessité d'infinies négociations par site, avec des enjeux locaux de partage qu'il ne semblait pas, lors de la création du HDH, raisonnablement possible de résoudre dans une perspective de trois à cinq ans. Cette approche fédérée stricte est néanmoins toujours promue par de nombreux CHU par exemple, qui ne veulent pas perdre le contrôle total de l'accès aux données de leur EDS.

En effet, lorsque des données sont hébergées au sein du catalogue du SNDS, un accès par projet peut être accordé par la CNIL et le CESREES sans autorisation préalable de la structure à l'origine de la base, si les critères de sécurité et d'intérêt public sont réunis. Ce principe semble juste car l'ensemble de ces données dispersées sur le territoire sont collectées et hébergées en très grande majorité avec de l'argent public ; mais la perte de contrôle sur l'accès, même très partielle, est source de crispations, souvent légitimes ou au moins compréhensibles, compte tenu des contraintes locales, notamment budgétaires<sup>45</sup>.

.....

45 Les considérations budgétaires liées à la pérennisation des EDS, mais également des registres et des cohortes, amènent à la question de la valorisation des données. Le sujet est épineux et clivant sur les plans conceptuel, juridique et pratique, mais dans une situation idéale, la « taxe » pour accéder aux données devrait être symbolique pour les projets d'intérêt général (notion déterminée par le CESREES), avec en arrière-plan un soutien étatique solide vers les structures elles-mêmes (EDS, registres, cohortes) afin de leur permettre d'échapper à la spirale néfaste de la construction de plans de financement, souvent très difficiles à mettre en place. Tout comme le SNDS, l'ensemble des structures hébergeant des données de santé de qualité devraient être considérées comme des structures d'intérêt public et financées en conséquence. La résolution de cette question financière semble être le levier le plus efficace pour lever les freins actuels au partage des données.

La seconde polémique concerne les modalités d'hébergement du nuage du HDH. Le choix d'Azure en 2019 était fondé sur un état des lieux qui positionnait Microsoft comme l'acteur le plus solide sur le plan technologique et sécuritaire<sup>46</sup>. Depuis, ce choix a été discuté au niveau politique à de nombreuses reprises, sur la base de la nécessité du maintien de la souveraineté numérique<sup>47</sup>.

### **Héberger : qu'est-ce que la souveraineté ?**

La souveraineté numérique est une notion problématique, aussi bien dans sa définition que dans son concept. Dans le domaine des nuages, pour résoudre la question de manière partielle mais formelle, deux labels de garantie français ont été créés ces dix dernières années :

- ▶ un premier référentiel a été lancé en 2018 par l'Agence du numérique en santé et révisé en 2023 : Hébergeur de Données de Santé (HDS)<sup>48</sup>, avec un cahier des charges qui assure le caractère national ou européen de la plupart des composantes de l'hébergeur ;
- ▶ un second référentiel, élaboré en 2016 par l'Agence nationale de la sécurité des systèmes d'information (ANSSI), SecNumCloud, et dont la dernière version (v. 3.2) datant de 2022<sup>49</sup>, propose un ensemble de règles de sécurité reflétant une bonne hygiène informatique et la protection des données en conformité avec le droit européen.

Dans la révision de 2023 du référentiel HDS, il est notamment nécessaire que « l'hébergement physique des données de santé [soit]

46 Voir par exemple les échanges à ce sujet sur le site associatif Ma Dada (<https://madada.fr>)

47 Voir exemple de questionnement politique sur les choix d'hébergement du HDH sur : <https://questions.assemblee-nationale.fr>

48 Détails de la révision 2023 du référentiel HDS : <https://technical-regulation-information-system.ec.europa.eu>

49 Dernière version 3.2 du référentiel SedNumCloud de l'ANSSI (<https://cyber.gouv.fr>)

réalisé exclusivement sur le territoire d'un pays situé au sein de l'Espace Economique Européen ». Un nuage à la fois HDS et Sec-NumCloud français présenterait ainsi des garanties combinées de souveraineté et de sécurité. Plus précisément, en termes de sécurité, même si le débat juridique reste très technique, les risques réels d'un hébergement de type Azure – non souverain – pour des données de santé chiffrées de la population française sont très faibles<sup>50</sup>, bien plus faibles dans les faits que les risques de piratage des EDS ou des systèmes d'information hospitaliers, qui peuvent contenir, à la différence du HDH, des données directement identifiantes des patients<sup>51</sup>.

**“ En termes de sécurité, même si le débat juridique reste très technique, les risques réels d'un hébergement de type Azure – non souverain – pour des données de santé chiffrées de la population française sont très faibles. ”**

Dans ce contexte de demande politique et populaire d'hébergement souverain, le HDH a pris acte de la nécessité d'une migration vers une solution dont le bénéfice principal réel serait une plus claire acceptation publique de l'institution, en travaillant sur des études de réversibilité vers d'autres nuages<sup>52</sup>, comme suggéré par un récent rapport interministériel intitulé « Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire

travaillant sur des études de réversibilité vers d'autres nuages<sup>52</sup>, comme suggéré par un récent rapport interministériel intitulé « Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire

50 Le cœur des interrogations sur les dangers de l'hébergement Azure est le *Cloud Act* américain (<https://www.congress.gov>) dont l'application ne peut être formellement exclue mais qui donnerait accès, s'il était utilisé par les Etats-Unis, à des données chiffrées présentant un intérêt très faible qui ne justifierait pas les efforts nécessaires à leur déchiffrement.

51 Les mesures de sécurité organisationnelles et techniques du HDH sont maximales : chiffrement des données stockées, segmentation des droits d'opération avec opération exclusive de la plateforme par les opérateurs du HDH, gestion sécurisée des comptes et des permissions, mise à disposition d'un espace de travail sécurisé, utilisation de briques techniques de sécurité, analyse des traces avec enregistrement de toutes les actions réalisées sur la plateforme, hébergement en Union Européenne dans les centres de données Microsoft des Pays-Bas certifiés HDS.

52 Discussion sur les études de réversibilité du HDH (<https://madada.fr>)

des données de santé »<sup>53</sup>. Dans l'intervalle précédant ce transfert, la possibilité de travailler avec l'hébergeur actuel a récemment été confirmée par la CNIL pour une durée de trois ans ; la CNIL a souligné l'importance d'un projet de migration, tout en validant le caractère sécurisé d'Azure, dans le cadre d'un projet porté par l'Agence européenne du médicament (EMA) et impliquant notamment des données du SNDS<sup>54</sup>. Le changement d'hébergeur sera ainsi très certainement effectué lorsque des initiatives considérées souveraines, comme le consortium Bleu<sup>55</sup>, seront opérationnelles. Alors que cette manœuvre marquera la fin des polémiques sur l'hébergement du HDH, elle ne répondra pas à la question de la réalité de la souveraineté numérique.

Cette notion peut ainsi être comprise en termes de restriction de périmètre, ce qui est son acception la plus fréquente, mais également, de manière positive, en termes de rayonnement. Le caractère symbolique, si ce n'est artificiel, de la version actuellement politiquement promue de la souveraineté transparaît dans la présence forte d'acteurs décriés, non souverains, comme Microsoft, au cœur même des solutions souveraines de demain comme Bleu, offrant certes des garanties de sécurité majeures (versant restrictif de la souveraineté), sans prise de conscience de la force que représente cette présence persistante des États-Unis en termes de stratégie d'influence et de constance dans la perte réelle de souveraineté. La souveraineté numérique de rayonnement, positive, consisterait à travailler de manière fédérée, en utilisant éventuellement des solutions non souveraines avec pragmatisme, pour contribuer activement à l'avancée de la science et au progrès de la médecine et

53 Rapport interministériel piloté par Jérôme Marchand-Arvier (<https://sante.gouv.fr>)

54 Rapport présenté à la CNIL concernant l'hébergement du projet EMC2 porté par l'EMA, et qui a débouché sur une autorisation : (1) étude d'hébergement : <https://esante.gouv.fr> ; (2) autorisation de la CNIL : <https://www.legifrance.gouv.fr>

55 Bleu est défini sur son site selon les termes : 'Bleu est une entreprise de services cloud, française, indépendante, fondée par Caggemini et Orange. Notre proposition : offrir la suite de collaboration et de productivité Microsoft 365, ainsi que les services Microsoft Azure, dans un cloud français sécurisé, ayant vocation à obtenir la qualification SecNumCloud 3.2 de l'ANSSI. Voir : <https://www.bleucloud.fr>

en positionnant dès aujourd'hui notre pays dans ce mouvement. Connaissant la richesse de son patrimoine de données, héritière de notre système de santé publique et centralisé, une France confiante et ouverte a un rôle majeur à jouer dans le futur de la médecine prédictive, sous réserve d'une transformation de notre conception de la souveraineté numérique.

### **Conclusion et perspectives**

La construction d'outils de médecine prédictive soulève des enjeux techniques, scientifiques et politiques. Tout comme le travail des physiciens des particules dépend de décisions politiques en raison des investissements matériels colossaux<sup>56</sup>, la prédiction rationnelle fondée sur des données massives est également sous le contrôle de décideurs dont les motivations ne sont pas systématiquement en adéquation avec les demandes de la science et de la médecine en marche. Cette incertitude institutionnelle ne minimise pas l'espoir majeur que représente le développement, dans un futur proche, de jumeaux numériques descriptifs et interactifs, qui seront les compagnons des praticiens dans le dialogue avec les patients, la préparation des plans de traitement, le suivi et la formation.

En chirurgie craniofaciale, les interventions sont lourdes et dangereuses, pratiquées chez des enfants très petits. La prédiction des histoires naturelles individuelles en fonction de marqueurs précoces est un défi majeur qui repose sur la construction de modèles agrégeant des quantités massives de données. L'application de ces modèles à la pédiatrie soulève la question de la croissance, qui complique la démarche prédictive, d'autant plus que les trajectoires de croissance sont elles-mêmes perturbées par les maladies associées

.....  
56 L'Organisation européenne pour la recherche nucléaire (ou CERN), le plus grand laboratoire de physique des particules au monde situé en Suisse près de Genève, disposait d'un budget de 1,405 milliards de francs suisses en 2022 (soit près de 1,5 milliard d'euros).

aux malformations craniofaciales<sup>57</sup>. Pour transformer la prise en charge de ces patients avec des outils numériques, les chirurgiens devront travailler avec des mathématiciens et des ingénieurs pour concevoir des outils prédictifs performants<sup>58</sup>, et bénéficier d'un soutien politique important pour lever les freins actuels pesant sur la réutilisation optimale des données de santé.

57 Par exemple, de nombreuses malformations craniofaciales surviennent dans le cadre de maladies osseuses congénitales qui perturbent l'ensemble des processus de croissance des tissus minéralisés (exemple : mutations activatrices des gènes *FGFR* causant les faciocraniosténoses et les ostéochondrodysplasies, associant malformations du crâne et anomalies osseuses autres des membres et/ou du rachis).

58 Rapport publié en 2023 par le Think Tank Arts et Métiers sur le rôle des ingénieurs dans le système de santé (<https://think-tank.arts-et-metiers.fr>)



