

A S S E M B L É E   N A T I O N A L E

X V <sup>e</sup>   L É G I S L A T U R E

# Compte rendu

## Office parlementaire d'évaluation des choix scientifiques et technologiques

- **Audition publique**, *ouverte à la presse*, sur le thème :  
Intelligence artificielle et données de santé : quelle collecte,  
quel accès aux données, pour quelles améliorations  
diagnostiques et thérapeutiques ? ..... 2

Jeudi 21 février 2019

Séance de 9 heures

Compte rendu n° 39

SESSION ORDINAIRE DE 2018-2019

**Présidence**  
**de M. Gérard Longuet,**  
***président***



## Office parlementaire d'évaluation des choix scientifiques et technologiques

Jeudi 21 février 2019

Présidence de M. Gérard Longuet, sénateur, président de l'Office

*La séance est ouverte à 9 heures*

**Audition publique, ouverte à la presse, sur le thème : Intelligence artificielle et données de santé : quelle collecte, quel accès aux données, pour quelles améliorations diagnostiques et thérapeutiques ?**

### OUVERTURE

**M. Gérard Longuet, sénateur, président de l'Office.-** Je voudrais, en qualité de président de l'Office, accueillir mes collègues parlementaires et souhaiter la bienvenue aux intervenants extérieurs qui participent à cette audition publique.

M. Cédric Villani, premier vice-président de l'Office, qui devait présider la première table ronde, est actuellement retenu par un rendez-vous avec madame la ministre de l'enseignement supérieur et de la recherche, mais nous rejoindra en milieu de matinée.

Le sujet qui nous réunit aujourd'hui, bien que déjà abordé par l'Office parlementaire, est très loin d'être épuisé, puisque l'on découvre chaque jour davantage sa complexité. Nous aurons, en tant que parlementaires, à le traiter dans le cadre des commissions et séances plénières de l'Assemblée nationale et du Sénat, par exemple prochainement dans le cadre de l'examen du projet de loi santé.

Cédric Villani a exploré la question de l'intelligence artificielle dans un rapport remis au premier ministre et au président de la République en mars 2018. Un chapitre complet y est dédié à la santé et s'articule autour des deux thématiques suivantes : l'intelligence artificielle au service d'une part de la médecine, d'autre part des politiques de santé. Ce travail faisait suite à un rapport réalisé par l'Office parlementaire et intitulé *Pour une intelligence artificielle maîtrisée, utile et démystifiée*. Nous avons aussi organisé une audition publique consacrée au sujet du dossier médical personnel.

Pour autant, le sujet est, comme je l'indiquais précédemment, loin d'être épuisé et se révèle toujours plus riche, important, profond et de plus en plus complexe au fur et à mesure que l'on avance.

Cédric Villani, dans son rapport au premier ministre, suggérait la constitution d'une plateforme, désormais inscrite dans le projet de loi relatif à l'organisation et à la transformation du système de santé déposé la semaine dernière par Mme Buzyn auprès de l'Assemblée nationale. L'audition que nous conduisons ce matin sera évidemment directement utile pour éclairer les travaux des commissions compétentes de l'Assemblée nationale, et ensuite je l'espère du Sénat.

Cédric Villani m'a demandé de vous rappeler le travail qu'il a conduit en relation avec Bernard Nordlinger, membre de l'Académie nationale de médecine, sur la thématique « santé et intelligence artificielle » et qui a fait l'objet, l'année dernière, d'une publication aux éditions du CNRS. Cet ouvrage regroupait 33 contributions, sur les aspects les plus divers de ce sujet.

Je voudrais remercier, au nom de l'Office parlementaire, l'ensemble des participants à cette audition, et notamment M. Nordlinger lui-même, présent dans la salle.

Cette réflexion s'inscrit évidemment dans une perspective de progrès de la connaissance et de l'action médicale par la science, mais pose la question de l'impact des données et des problèmes de société engendrés par les données massives et l'accès à des informations concernant chacun d'entre nous et dont on imagine qu'elles doivent être respectées comme faisant partie de la personnalité de chaque individu, dans ce qu'elle a de plus intime. Cédric Villani concluait son rapport de mars 2018 par la phrase suivante : « *Il n'est pas question de remplacer le médecin par la machine. L'enjeu est bien d'organiser les interactions vertueuses entre l'expertise humaine et les apports de l'intelligence artificielle, dans l'exercice quotidien de la médecine* ».

Les travaux d'un organisme comme l'Office, qui s'efforce d'éclairer les parlementaires sur les avancées de la science dans divers domaines, ont aussi pour vocation de faire en sorte que la France se dote d'une politique nationale, la question étant en l'occurrence de savoir si cette intelligence artificielle utilisée au service de la santé et des politiques de santé peut constituer un axe majeur de rayonnement mondial de la France, en lui donnant un avantage stratégique par une connaissance avancée dans ce secteur. Ce rayonnement mondial est-il réel ? Où la France se situe-t-elle exactement ? C'est là l'un des aspects dont il nous faudra débattre ce matin.

La première des deux tables rondes de cette audition concerne la question de la collecte et de l'accès aux données de santé. Je rappelle à ce sujet que l'article 11 du projet de loi relatif à l'organisation et à la transformation du système de santé énonce des propositions de règles assez précises en la matière.

La seconde table ronde portera sur les finalités de cette collecte, au service des soins, des diagnostics, des thérapeutiques.

Nous allons accueillir des intervenants d'horizons très divers : c'est là tout l'intérêt d'une telle réunion que de favoriser l'échange des points de vue, dont chacun mérite d'être entendu. J'insiste toutefois sur la nécessité de respecter les temps de parole prévus.

## **PREMIÈRE TABLE RONDE : LA COLLECTE DES DONNÉES : NATURE DES DONNÉES, MODALITÉS DE COLLECTE, ENJEUX JURIDIQUES ET PROTECTION DES DONNÉES**

**Présidence de M. Gérard Longuet, sénateur, président de l'Office, puis de M. Cédric Villani, député, premier vice-président.**

**M. Gérard Longuet, sénateur, président de l'Office.-** Cette première table ronde va nous permettre d'entendre des interventions de l'administration centrale, par l'intermédiaire de la direction de la recherche, des études, de l'évaluation et de la statistique (DREES), mais aussi de la Caisse nationale d'assurance maladie (CNAM), d'un médecin spécialiste de radiologie et d'imagerie médicale et d'un représentant de la Commission nationale informatique et libertés (CNIL) travaillant sur la régulation et la protection de l'accès aux données. Nous aurons également le plaisir d'accueillir un membre du Barreau, un spécialiste de la réflexion éthique dans le domaine de l'intelligence artificielle en lien avec la santé et des chercheurs en algorithmie appliquée au domaine de la santé, tant dans le secteur public que privé. L'Office a également mobilisé, à l'occasion de cette audition, des représentants d'industriels et de *start-up*, qui nous feront partager leur enthousiasme face à cette déclinaison de l'intelligence artificielle et sa maîtrise au service de la santé. Nous accueillerons enfin un représentant des associations de patients : il m'apparaît en effet très important que les témoignages issus des expériences et parfois des souffrances capitalisées par les patients puissent s'exprimer au travers d'une démarche construite et pas seulement mue par l'émotion. Ceci constitue une véritable richesse pour le débat.

Toutes ces contributions nous seront d'une très grande utilité et je remercie une fois encore l'ensemble des intervenants du temps qu'ils ont accepté de consacrer à ce moment de vie parlementaire et, ce faisant, à la vie publique de notre pays.

Je signale par ailleurs que notre collègue Valeria Faure-Muntian, députée de la Loire, très impliquée dans la réflexion sur l'intelligence artificielle et les sujets autour du numérique, analysera les questions des internautes.

**Mme Valéria Faure-Muntian, députée.-** L'intelligence artificielle est un sujet d'actualité et d'avenir, sur lequel nos collègues se sont penchés dès 2017. Nous avons par ailleurs, avec le rapport de Cédric Villani, auquel j'ai eu la chance de contribuer, essayé de balayer l'ensemble des enjeux et des sujets, parmi lesquels la santé s'est imposée comme un domaine où la France pouvait, du fait notamment des nombreuses entreprises investies dans ce secteur, être un moteur d'excellence, une locomotive à l'échelle européenne.

Il est très intéressant pour nous d'entendre les experts, que je remercie d'être présents ce matin, pour envisager la manière de gérer ces développements, dans le but d'apporter des services complémentaires à nos concitoyens. Le

président de la République a souligné la nécessité que la technologie soit au service de l'humanité. La question se pose donc aujourd'hui de savoir comment l'intelligence artificielle pourra être mise au service de la santé de nos concitoyens.

**M. Gérard Longuet, sénateur, président de l'Office.-** Nous allons commencer cette table ronde en accueillant M. Jean-Marc Aubert, directeur de la recherche, des études, de l'évaluation et de la statistique (DREES) au ministère des solidarités et de la santé et Mme Stéphanie Combes, cheffe du projet *Health data hub* au sein de cette direction. Je rappelle que l'article 11 du projet de loi relatif à l'organisation et à la transformation du système de santé définit le statut et l'organisation de cette plateforme.

### **Interventions**

**M. Jean-Marc Aubert, directeur de la recherche, des études, de l'évaluation et de la statistique (DREES).-** Voici quelques mois, suite à la feuille de route rédigée par un groupe de travail confié à Mme Polton, MM. Wainrib et Cuggia, la ministre de la santé m'a confié la responsabilité d'essayer de mettre en œuvre la plateforme des données de santé dont le rapport du député Cédric Villani proposait la construction.

Les outils d'intelligence artificielle se développent depuis plusieurs années dans le domaine de la santé et commencent, dans un certain nombre de pays dont la France, à rendre des services. On pense par exemple aux outils apportant une aide au diagnostic en imagerie ou en matière de chirurgie, avec des robots utilisant des algorithmes issus de ces recherches.

Si la France souhaite un jour occuper une position de *leader* dans le domaine, elle dispose déjà de certains atouts, dont de fortes compétences scientifiques. Il manque en revanche une infrastructure. Le XIX<sup>e</sup> siècle s'est construit autour du chemin de fer ; le XX<sup>e</sup> siècle a vu le développement des autoroutes. Aujourd'hui, l'infrastructure majeure réside dans le numérique et la capacité de collecte, de stockage et d'exploitation des données qui vont permettre le développement de nouveaux services.

Dans le domaine de la santé, la France dispose aujourd'hui d'une base médico-administrative qui la place en pointe, mais pâtit, par rapport aux pays les plus avancés, d'un retard en termes de capacité à mêler ces données médico-administratives à d'autres données. Dans certains pays nordiques par exemple, des études sont menées assez facilement en vie réelle sur des cohortes de plusieurs millions de patients et mêlent des données de soin, des données cliniques et, demain, des données génétiques.

Il est important que la France puisse se doter d'une telle infrastructure. C'est la raison pour laquelle le projet de loi déposé sur le bureau de l'Assemblée

nationale prévoit dans son article 11 non seulement la constitution de la plateforme, sur laquelle reviendra Stéphanie Combes, mais aussi et surtout le fait que toutes les données cliniques recueillies au cours d'un acte ou d'une intervention remboursés par l'assurance maladie pourraient faire partie, sous forme évidemment anonymisée, d'un système de données ouvert, permettant notamment des recherches scientifiques, publiques ou privées, en intelligence artificielle. En termes de soin, les professionnels de santé auront évidemment accès à ces données, comme ils le font déjà aujourd'hui. Il s'agit d'un point important, puisque ce principe s'imposera à l'ensemble des acteurs et producteurs de soin.

Cette démarche est finalement assez proche de celle adoptée dans le domaine de la recherche publique, avec des initiatives comme « la science ouverte » : il s'agit de considérer que toutes les données qui aujourd'hui existent de façon disséminée, stockées dans des entrepôts hospitaliers, regroupées dans des cohortes, consignées dans des registres, n'ont pas vocation, dans la mesure où leur collecte est en général financée par la puissance publique, à être utilisées seulement par l'équipe de recherche de l'institut qui les a recueillies, mais à être mises à disposition de l'ensemble du monde de la recherche et, au-delà, des gestionnaires du système de santé et des *start-up* susceptibles de proposer des services nouveaux. Il s'agit de l'un des aspects importants du projet de loi, sur lequel la ministre des solidarités et de la santé est fortement engagée, considérant qu'avant même le *hub*, qui permettra ensuite de faciliter l'accès à l'ensemble de ces données, cet élément apportera à notre pays la capacité de développer l'intelligence artificielle, mais aussi des recherches et des services pour favoriser une meilleure prise en charge des patients.

Je vais laisser la parole à Stéphanie Combes, qui va vous présenter la plateforme de données.

**Mme Stéphanie Combes, cheffe du projet *Health data hub*, direction de la recherche, des études, de l'évaluation et de la statistique (DREES).**- La plateforme de données de santé ou *Health data hub* va venir remplacer l'Institut national des données de santé (INDS), avec pour mission d'accompagner les utilisateurs du SNDS (Système national des données de santé) dans leurs démarches et leurs demandes d'accès aux données.

Avec l'élargissement du SNDS que vient de présenter Jean-Marc Aubert, le périmètre est beaucoup plus vaste que précédemment. Il sera très intéressant pour les utilisateurs de données de bien comprendre de quoi est composé le patrimoine de données de santé français, ce que l'on peut en faire et comment réaliser des agrégations ou des rassemblements de jeux de données de santé. On estime en effet que les plus beaux cas d'usage viendront de la réunion de différentes sources de données. Ainsi, associer par exemple des données d'imagerie médicale et des données de parcours de soin pourrait permettre de raccourcir les temps de diagnostic, d'identifier des états précancéreux et de mettre en place des prises en

charge beaucoup plus rapidement. J'imagine que de nombreux autres exemples seront donnés, notamment lors de la deuxième table ronde.

Aujourd'hui, ce type de projet permettant d'associer différents jeux de données de santé peut prendre plusieurs années. L'idée de la plateforme de données est d'offrir un accompagnement et d'accélérer la réalisation de ces projets.

La deuxième mission de la plateforme sera de mettre à disposition les jeux de données de manière effective, dans un espace sécurisé, avec des outils à l'état de l'art. L'espoir est de voir se développer davantage de cas d'usage d'intelligence artificielle notamment. Ceci devrait faciliter l'innovation et permettre par exemple à des *start-up*, qui doivent aujourd'hui, pour effectuer de l'apprentissage d'algorithmes, récupérer des données à l'étranger, d'utiliser le patrimoine de données de santé français. L'objectif étant que les citoyens et patients français bénéficient des innovations qui en découleront, il est essentiel que les algorithmes soient entraînés sur des données nationales. Il est bien évident que la communauté d'utilisateurs du *Health data hub* va par ailleurs bien au-delà des seules *start-up* et englobe chercheurs, institutionnels et industriels.

Une autre mission de la plateforme consistera à accompagner les producteurs dans l'exercice de collecte des données, puisque les cas d'usage des innovations vont dépendre de notre capacité à alimenter et à mettre au bon niveau de qualité notre patrimoine de données. Or ceci requiert des ressources financières, techniques, ainsi que des compétences très pointues qui ne sont aujourd'hui pas toujours faciles à « capter ». Disposer d'un projet national et visible pourrait nous permettre d'être attractifs pour des talents du monde numérique, de l'informatique et de l'analyse de données.

Le *hub* sera également une sorte de carrefour, qui va pouvoir mettre en relation producteurs et utilisateurs. Il existe déjà en France de nombreuses et très intéressantes initiatives, qui mériteraient d'être mieux valorisées, mieux diffusées dans l'écosystème. Nous avons donc besoin d'un tel espace permettant aux acteurs de se parler et d'échanger. Cette plateforme sera aussi le moyen de faire le pont avec la société civile et d'être très transparent sur l'utilisation qui sera faite des données de santé, comme ceci est aujourd'hui prévu avec l'INDS.

Très concrètement, le *Health data hub* transformera l'INDS. Il s'agira d'une structure en tant que telle, avec ses équipes. Comme l'INDS, il est prévu qu'elle prenne la forme d'un groupement d'intérêt public. Ce choix résulte de la comparaison de différentes formes juridiques, allant de l'établissement public à la société par actions simplifiée. En termes de gouvernance, les parties prenantes de l'Institut national des données de santé auront leur place dans le futur *Health data hub*. Le périmètre s'élargissant, s'y ajouteront d'autres acteurs, par exemple des représentants de centres hospitaliers.



**M. Gérard Longuet, sénateur, président de l'Office.-** Je cède la présidence de cette table ronde à Cédric Villani, qui vient de nous rejoindre.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, cher Gérard, M. le président. Je vous prie d'excuser mon retard : je sors d'une réunion de lancement d'une mission qui sera consacrée à la programmation pluriannuelle de la recherche, organisée par la ministre de l'enseignement supérieur et de la recherche, Mme Frédérique Vidal, sur un sujet qui concerne aussi l'Office au premier plan. Nous aurons l'occasion d'en reparler.

Nous poursuivons cette première table ronde en donnant la parole à M. Claude Gissot, directeur de la stratégie, des études et de la statistique à la Caisse nationale d'assurance maladie (CNAM), direction en charge en particulier de la célèbre base de données au nom si poétique de SNIIRAM (système national d'information inter-régimes de l'Assurance maladie), plus important collecteur public français de données de santé.

Il ne s'agissait bien évidemment pas, dans la composition de cette table ronde, de donner la parole à tous les responsables des différentes bases de données constituant le SNDS, dont nous venons d'entendre qu'il allait être élargi dans le cadre du projet de loi du gouvernement ; mais le SNIIRAM en est sans doute l'élément le plus emblématique.

**M. Claude Gissot, directeur de la direction de la stratégie, des études et de la statistique à la Caisse nationale d'assurance maladie (CNAM).-** Je vais revenir sur la constitution du SNIIRAM et les données qu'il contient.

Cette histoire, qui place aujourd'hui la France en pointe sur la question de la collecte des données médico-administratives, a commencé dans les années 2000. Le SNIIRAM s'est ensuite développé en différentes étapes et enrichi au fil du temps. Comme l'a souligné M. Aubert, les principales données le composant sont recueillies au moment de la liquidation du remboursement des soins. Il faut savoir qu'environ 1,2 milliard de feuilles de soin sont traitées chaque année par l'Assurance maladie et des dizaines de millions de séjours hospitaliers intégrés : cet ensemble de données a constitué le SNIIRAM, première brique essentielle dans le dispositif.

On parle de « données médico-administratives » dans le sens où il ne s'agit pas de données médicales recueillies par le professionnel de santé lui-même : ce système ne comporte pas de diagnostics médicaux directs, ni de résultats d'analyses. Tous ces éléments sont absents du SNIIRAM, ce qui constitue l'une de ses limites, mais nous projette aussi dans l'avenir, dans les recueils que nous devons organiser pour compléter l'existant et développer ainsi les usages du point de vue de la connaissance, de la recherche, de l'appui à la décision médicale, etc.

Concrètement, le traitement de plus d'un milliard de feuilles de soins, qui représentent des centaines de téraoctets de données, nécessite des infrastructures adaptées. Nous avons actuellement près d'une quinzaine d'années d'ancienneté dans les données collectées, ce qui en fait un gisement extrêmement important, propice au développement et à l'utilisation de technologies différentes de celles que nous connaissons aujourd'hui, en termes notamment d'intelligence artificielle et d'algorithmie.

Le SNIIRAM recense par exemple tous les médicaments délivrés par les pharmacies, tous les actes médicaux réalisés par les médecins et les professionnels de santé, avec un degré de détail très important, dans la mesure où les nomenclatures ont été affinées au fil du temps. Ceci a permis d'atteindre un degré de précision extrêmement important dans ces domaines, qui conduit finalement à la capacité d'en créer les usages. Il faut savoir que les données sont rattachables à un individu précis, dans un environnement pseudonymisé : ainsi, les nom, prénom et adresse des personnes n'existent pas dans la base et les identifiants individuels sont des pseudonymes. Bien entendu, le degré de précision de ces données (dates de soin, actes pratiqués, professionnels impliqués, établissements ayant accueilli les patients, etc.) fait que leur croisement les rend indirectement identifiantes. Il est donc essentiel de les protéger. C'est la raison pour laquelle les procédures d'accès aux données sont strictement encadrées par des dispositifs validés avec la CNIL en termes de procédures et assurant fondamentalement la protection de la vie privée des gens qui confient ces données à la puissance publique par l'intermédiaire des producteurs de soin. Ces dispositifs permettent à chaque personne autorisée, selon ses besoins et la finalité de ses requêtes, d'accéder à certaines informations, dans un cadre général visant à la protection des données.

La CNAM, en tant que régulateur, a beaucoup utilisé ces données depuis quinze ans. Nous avons été les premiers à en démontrer l'utilité, à travers un certain nombre d'études, dont quelques-unes sont restées célèbres, concernant notamment l'aspect délétère de divers médicaments. Les travaux récents et assez médiatisés menés sur des médicaments comme le valproate ou l'Androcur sont permis par cette base et en révèlent l'utilité.

Nous avons également développé au cours des dernières années la détermination indirecte des pathologies au travers du contenu de la base. On peut en effet déduire des médicaments pris par une personne la pathologie dont elle souffre. Ce n'est pas un diagnostic porté par les médecins, mais une déduction effectuée à partir de la combinatoire des informations contenues. Ce point constitue la première étape vers une utilisation algorithmique. Aujourd'hui, l'usage que nous faisons de ces données est essentiellement déterministe, mais il ouvre la voie à d'autres opérations plus complexes, utilisant des technologies plus avancées.

Je terminerai cet exposé par un volet plus prospectif. Ce socle de données médico-administratives sera d'autant plus intéressant qu'il sera potentialisé par le reste des données. Il a été question des résultats de biologie,

d'éléments d'imagerie, d'autres données encore : le fait de mettre toutes ces informations ensemble permettra aux usages de se développer. Il faut également envisager l'utilisation des données produites par le patient lui-même : tous les éléments qu'il pourra reporter sur son état de santé ou les résultats vécus dans le cadre de son processus de soin sont de nature à améliorer la compréhension de l'ensemble de la prise en charge, donc au final à apporter des arguments et des outils à la politique publique de santé.

Nous sommes une brique de base et avons devant nous tout un pan à développer, du point de vue des données et de leur recueil, mais aussi en matière d'investissement dans la technologie. Il faut notamment mettre les outils d'exploitation à l'état de l'art, afin d'améliorer l'efficacité d'exploitation des données. Ceci suppose notamment d'investir dans les ressources humaines, c'est-à-dire dans les compétences techniques et métiers, avec des professionnels connaissant parfaitement le système de santé et les prises en charge, mais aussi des *datascientists*, dans le prolongement de l'action des statisticiens.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci, M. Gissot. Pourriez-vous nous préciser si vous êtes satisfait des formats de collecte de données tels qu'ils existent actuellement et de la qualité des supports technologiques sur lesquels ils sont disponibles ?

Vous avez par ailleurs évoqué des besoins en ressources humaines, en investissements : avez-vous une idée du chiffrage permettant de satisfaire vos rêves en la matière ?

Vous n'avez en outre pas abordé la question de la cyberprotection : où en êtes-vous dans ce domaine ?

**M. Claude Gissot.-** Nous disposons aujourd'hui d'un système d'exploitation des données qui date des années 2000. Nos bases de données sont très proches de la production et demandent un travail très important pour construire les sous-ensembles de données utiles ensuite aux chercheurs. L'accumulation des données et la nécessité de disposer d'un système plus agile et facile à manipuler nous a ainsi poussés à développer un partenariat avec l'École polytechnique.

En matière de cyberprotection, il faut savoir qu'un référentiel de sécurité a été produit voici deux ans, qui doit être totalement opérationnel d'ici fin mars 2019. Il s'agit d'un élément extrêmement important, l'ouvrage étant sans cesse remis sur le métier afin d'offrir toutes les garanties de sécurité nécessaires.

Il est difficile d'avancer des coûts chiffrés. Nous avons évalué le coût du SNDS, indépendamment de l'élargissement actuel tel que présenté par la DREES. Le montant s'élevait à plusieurs dizaines de millions d'euros par an en coûts complets, incluant les infrastructures, la maintenance, la mise en œuvre, les projets associés, etc. Cette somme peut paraître considérable, mais prend une autre

dimension si on la compare à la charge globale de 200 milliards d'euros que constituent les dépenses socialisées en matière de santé. Eu égard aux enjeux, on peut estimer que ceci mériterait certainement un investissement important de la part de la collectivité.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, M. Gissot.

Nous continuons notre tour de table sur cette première thématique relative à l'organisation de la collecte des données avec M. David Gruson, membre du comité de direction de la chaire santé de l'Institut d'études politiques de Paris et fondateur d'Ethik-IA. Nous souhaiterions notamment savoir comment vous évaluez le rapport du Comité consultatif national d'éthique sur le thème « IA et numérique en santé » et connaître les propositions d'Ethik-IA sur la question de la régulation positive, normative, de l'intelligence artificielle appliquée à la santé, afin de garder ce secteur sous contrôle, sans être trop pesant. Quelles méthodologies, quelles modalités pour les métiers de la santé ?

**M. David Gruson, membre du comité de direction de la chaire santé de l'Institut d'études politiques de Paris, fondateur d'Ethik-IA.-** Le rapport du Comité consultatif national d'éthique auquel vous faites référence fait suite à la mission que j'ai eu l'honneur de co-diriger avec Claude Kirchner, président du conseil d'éthique d'Inria. Comme l'ont montré de multiples travaux, dont l'ouvrage que vous avez, M. Villani, coordonné avec Bernard Nordlinger ici présent, l'IA se développe massivement dans le système de santé, avec un phénomène d'espace temps et de rapidité de déploiement qui doit amener à prendre des orientations en termes de régulation éthique, en essayant de soutenir l'innovation tout en maîtrisant les risques. C'est cette quadrature du cercle que le CCNE a tenté d'aborder dans le rapport rendu public en novembre 2018. Cette réflexion représente sans doute un point de bascule. Le Comité d'éthique structure dans ce document une position que je vais résumer très synthétiquement, autour d'une pyramide de risques à trois étages, du plus important au moins significatif.

Le CCNE considère que le risque majeur serait de ne pas s'ouvrir à l'IA, au numérique, au pilotage par les données. J'ai fait, en tant que directeur de CHU, l'expérience de ce que l'insuffisance du pilotage par les données génère comme situations de non qualité et de sous-efficience, lesquelles sont profondément non éthiques. Sans doute allez-vous, au fil de cette audition, entendre divers témoignages en ce sens. Il s'agit d'un sujet que j'ai abordé avec Gérard Raymond sur le champ du diabète ou encore avec Olivier Véran, que je salue, sur la question de l'insuffisance rénale chronique. Il n'est pas éthique, en résumé, de bloquer la diffusion de l'IA.

Le deuxième risque identifié tient au fait que la médecine algorithmique se diffuse très vite ailleurs et que le numérique rend nos systèmes de santé de plus en plus poreux entre eux. Si nous ajoutons des strates de sur-réglementation en

France, ceci encouragera en pratique les professionnels et les patients à recourir à des instruments de médecine algorithmique ailleurs, dont on ne pourra pas garantir le principe éthique et notamment les modalités de recueil du consentement dans la collecte des données.

Une fois posés ces deux préalables fondamentaux, qui traduisent sans doute une bascule d'idée sur ce sujet, on peut aborder les risques éthiques intrinsèques à l'intelligence artificielle, autour d'une série de propositions pour la prochaine révision de la loi relative à la bioéthique. Nous avons d'ailleurs constaté avec satisfaction que certains de ces éléments avaient été relayés dans le rapport de la mission d'information sur la bioéthique rapportée par M. Jean-Louis Touraine et portés au Sénat par le président de sa commission des affaires sociales, M. Alain Milon.

La première proposition concerne un élargissement du devoir d'information du médecin au patient, afin que ce dernier soit informé de ce que la proposition thérapeutique qui lui est faite se fonde sur les résultats produits par un système d'intelligence artificielle.

La deuxième modification suggérée est celle d'une « garantie humaine » de l'intelligence artificielle : ceci renvoie à l'idée d'une supervision de l'IA non pas à chaque étape (ce qui en bloquerait la capacité d'innovation), mais dans une logique de supervision pratique. Nous avons par exemple, avec la Société française de télémedecine, proposé l'idée d'une télémedecine de garantie humaine : ceci est très intuitif et consiste, pour un médecin de première ligne auquel l'IA propose un diagnostic sur lequel il a un doute par rapport au cas clinique, à solliciter un deuxième avis humain, sous forme d'une téléexpertise sur la proposition algorithmique. Ce type de procédure entre tout à fait dans le droit actuel de la téléexpertise et dans le régime financier fixé pour ce domaine par la convention médicale de septembre dernier. Ces dispositifs de pilotage humain pourraient également consister en la mise en place de collèges de garantie humaine associant médecins, soignants, représentants des usagers, qui assureraient le suivi de l'évolution de l'algorithme au fil du temps et vérifieraient que ce dernier reste efficace médicalement et responsable éthiquement. On peut imaginer développer, autour de ce cœur de proposition, un écosystème de régulation positive, sous la forme, comme le proposait le Conseil de l'ordre des médecins, de *soft law*, de régulation souple, de *self-compliance*, d'outils de bonnes pratiques, le cas échéant sous le pilotage de la Haute autorité de santé.

Je considère, à titre personnel, qu'il conviendrait sans doute de compléter ces propositions, d'un point de vue pratique, par deux lignes de recommandations complémentaires. Sur le terrain du droit, le CCNE relevait dans son rapport le caractère très positif de la mise en place d'une plateforme sécurisée de données telle que le *Health data hub*. Il s'agit évidemment d'une avancée. Je dirais pour ma part que si l'on voulait compléter le dispositif pour permettre au *hub* d'atteindre sa pleine efficacité dans un contexte caractérisé par le fait que l'innovation se

développe très vite ailleurs, se poserait la question pratique consistant à permettre un usage assez rapide des données déjà collectées. A partir du moment où le *hub* se met en place sous l'égide des pouvoirs publics et est entouré de principes juridiques et éthiques constitutifs, pourquoi ne pas prévoir, pour les données déjà collectées, un principe permettant, à condition que le traitement soit d'intérêt légitime, public, que s'applique le consentement présumé ? Ceci permettrait assurément un gain de temps dans le développement d'outils de médecine algorithmiques dans le cadre protecteur qui est le nôtre, avec le cas échéant une capacité d'opposition. Certains régimes de ce type existent dans le champ du traitement des données, mais aussi de la santé, pour des actes aussi graves que les prélèvements d'organes.

Ma deuxième proposition a à voir avec une éthique entendue au sens large, dans l'optique d'une régulation positive de l'accompagnement. Toutes ces évolutions ont évidemment un impact important pour les professionnels de santé. On entend par exemple régulièrement dire dans le débat public que la radiologie est une spécialité en voie de disparition du fait de l'IA. Les analyses que nous avons pu mobiliser, notamment dans le cadre d'un rapport rédigé pour l'Institut Montaigne, montrent en fait que le point d'impact ne porte pas à court terme sur les spécialités médicales et soignantes. Il faut travailler sur l'évolution de la formation médicale, réfléchir aux effectifs qui seront nécessaires. La Conférence des doyens des facultés de médecine a proposé d'insérer un module de formation des étudiants en médecine à l'IA dès la première année de 2019. Sans doute faut-il également regarder ailleurs : le sujet principal est selon moi celui des fonctions supports que sont l'administration, la gestion, le pilotage des fonctions logistiques et médico-techniques. Des scénarios de chiffrage ont été établis, qui montrent que les emplois automatisables dans le système de santé sur ces fonctions sont de l'ordre de 40 000 équivalents temps plein en fourchette basse et 80 000 en fourchette haute. Le potentiel d'automatisation est donc considérable et doit évidemment être mis en perspective avec les 1,2 million d'emplois du secteur. Ceci implique le déploiement, dans les meilleurs délais, d'un plan de gestion prévisionnelle des emplois et des compétences, afin d'essayer d'adapter ces profils aux besoins du système de santé de demain. Sans doute est-ce le point essentiel sur lequel l'effort doit porter en matière d'accompagnement du déploiement de l'IA en santé : ceci participe de principes de régulation éthique.

**M. Cédric Villani, député, premier vice-président de l'Office.** Merci beaucoup, M. Gruson. Nous retenons en particulier votre appel à une ouverture et un décloisonnement, considérant que le principal danger éthique serait de ne pas accueillir et piloter le changement. Vous insistez également sur les fonctions supports, pour lesquelles la question de l'automatisation doit être regardée de près et en urgence.

Nous poursuivons cette audition avec Mme Isabelle Zablit-Schmitz, coprésidente du comité Santé et administratrice de Syntec Numérique. Nous savons que Syntec Numérique, premier syndicat professionnel de l'écosystème numérique

français, est un acteur majeur en tant que relais d'opinion, force de proposition, d'organisation et de structuration de toute la communauté numérique. Il regroupe des entreprises de services du numérique, des sociétés de conseil en technologie, des éditeurs de logiciels, des acteurs du *web* et comprend un comité Santé qui suit précisément les sujets d'intérêt de ce matin, avec un groupe de travail « données et IA ». En bref, vous avez déjà travaillé sur la question qui nous réunit aujourd'hui et dégagé une position commune.

J'ajoute que vous êtes, à titre personnel, cofondatrice de Wellfundr, *start-up* de l'e-santé. Vous êtes donc particulièrement qualifiée, tant du fait de votre parcours que des institutions que vous représentez, pour nous faire part de la position de l'écosystème numérique. Votre position commune s'intitule *Développer un écosystème pérenne de valorisation des données et d'innovation*, ce qui fait parfaitement écho à notre état d'esprit.

**Mme Isabelle Zablit-Schmitz, coprésidente du comité Santé et administratrice de Syntec Numérique.-** Syntec Numérique s'est effectivement engagé de longue date, par l'intermédiaire de son comité Santé, sur ces questions. Je souhaiterais simplement compléter la présentation que vous en avez faite, et dont je vous remercie, en précisant que nous représentons aujourd'hui environ 80 % de l'écosystème du numérique, qui correspond à quelque 500 000 emplois en France. Nous comprenons, comme vous l'avez indiqué, des organisations différentes, puisque nous réunissons à la fois une trentaine de grands groupes, très médiatiques, mais aussi, à l'autre bout du spectre, 850 *start-up*. En tant qu'écosystème du numérique, nous ne sommes par ailleurs pas essentiellement consacrés à la santé. Toutes les entreprises adhérentes sont présentes en France, mais certaines d'entre elles sont également internationales.

Il me semble important de souligner ces différents aspects dans la mesure où ils nous permettent d'avoir une vision sur la santé émanant de spécialistes, mais aussi de généralistes s'adressant à l'ensemble des secteurs, la particularité de nos entreprises étant justement d'aider les autres entreprises et secteurs de l'économie à évoluer sur la voie du numérique et du progrès.

Pourquoi avoir mis en place ce comité Santé ? Nous avons estimé, voici une dizaine d'années, qu'il était nécessaire, à partir de la connaissance des différents adhérents composant notre organisation, de nous pencher plus précisément sur le champ de la santé, car nous percevions déjà tout le potentiel du numérique dans ce domaine, à une époque où il n'était pas encore considéré comme un vecteur de transformation du secteur. Nous ne pouvons par conséquent que saluer le virage pris récemment, consistant à placer le numérique au cœur de la stratégie de transformation du système de santé.

Le numérique conduit tout naturellement à aborder la question de la gestion des données. Gérer les données est véritablement un métier. Lorsqu'il s'agit de données sensibles telles que les données de santé, on comprend immédiatement

l'importance, comme l'a souligné David Gruson, d'élaborer un cadre éthique et de sécurité. J'y reviendrai.

Quelque 150 entreprises, généralistes et spécialistes, contribuent aux travaux de notre comité Santé sur l'IA et la donnée de santé, mais aussi sur le sujet des plateformes et des infrastructures numériques, ainsi que des objets connectés. Bien qu'il s'agisse de groupes de travail différents, tous sont liés par le sujet commun de la gestion des données.

Syntec Numérique a par ailleurs lancé l'an dernier un grand débat sur la donnée, que vous avez pu suivre, et travaille également, en commissions, sur le sujet de l'éthique et de la cybersécurité.

Nous considérons que la gestion de la donnée est aujourd'hui un véritable métier. Nous nous positionnons ainsi depuis plus de dix ans en co-construction de ce que ceci implique sur le terrain. Nous avons ainsi salué et contribué à l'émergence du projet *Health data hub*, dans le prolongement de la mission conduite par M. Villani sur l'intelligence artificielle.

Nous observons qu'il est essentiel de structurer une filière des données de santé en tant que telle. Nous employons à dessein le terme de « filière » dans la mesure où nous considérons que la gestion des données concerne non pas un, mais des métiers. Ce sont pour la plupart des métiers nouveaux. Plusieurs intervenants précédents ont rappelé l'importance de la formation et des compétences. Il s'agit d'un sujet d'attention, voire de préoccupation, que nous souhaitons également porter. Nous constatons en effet, chez certains de nos adhérents, un manque cruel de compétences, assez paradoxal dans le paysage économique actuel caractérisé par un virage numérique global. Nous souhaitons ainsi relayer avec force cet appel à la formation. Syntec Numérique s'est investi de longue date dans ce domaine, avec différentes actions. Il ne s'agit pas uniquement de former des professionnels des technologies, mais aussi d'acculturer les personnes dont le métier est appelé à se transformer pour s'adapter à ces nouveaux modes de fonctionnement. Quelque 10 000 offres d'emplois restent chaque année à pourvoir et cette situation est appelée à empirer au fil du temps si l'on n'y prend garde. Il s'agit donc d'un point d'attention absolument essentiel.

Nous souhaitons aussi que cette filière des données de santé soit en phase avec l'évolution de l'économie et du potentiel d'innovation autour de la *data* médicale. Ceci signifie qu'il faut veiller à ce que ces données circulent. Le *Health data hub* constitue une première brique. Citons également l'Espace numérique de santé, autre projet inclus dans le plan « Ma santé 2022 ». Ce sont des éléments essentiels pour que les données de santé soient une source d'amélioration au service des patients et des professionnels de santé. Il s'agit à présent d'organiser ces travaux, de les sécuriser, mais aussi de s'attacher à valoriser les données de santé.



Ces différents axes appellent à chaque fois des initiatives et des compétences.

En ce qui concerne la sécurisation des données, nous avons accompagné un sujet sur lequel la France est précurseur, à savoir la sécurisation de l'hébergement des données de santé. La France a innové dans ce domaine, avant même l'arrivée du règlement général européen sur la protection des données. Nous avons salué l'intérêt de cette démarche, tout comme le passage de cette réglementation à une certification basée sur des standards internationaux. En effet, lorsqu'il s'agit de créer une filière de données de santé, il faut être vigilant et créer un cadre permettant la fluidité de l'information en toute sécurité, mais aussi ne pas se cantonner à un cadre français et s'appuyer sur des standards internationaux. Cette dimension est essentielle.

Il est également très important de sécuriser les systèmes d'information et le socle technologique de base. Si l'on veut déployer une filière de données de santé, il faut évidemment que ces données s'appuient sur un socle technologique solide. Ceci a constitué un sujet d'inquiétude de longue date dans les travaux du comité Santé, notamment au vu du sous-investissement historique réalisé en France dans ce domaine, ce qui est finalement assez logique puisque le numérique n'était jusqu'alors pas considéré comme un vecteur de transformation.

Nous pensons qu'il est nécessaire de déployer des outils technologiques pour veiller à la circulation de ces données. Il est question également d'authentification, d'accessibilité, d'hébergement. Tout ceci fait appel à des compétences et à des outils technologiques, à la construction desquels nous sommes disposés à contribuer.

Nous accordons en outre une grande importance au respect des droits de la personne. Syntec Numérique et le CIGREF ont ainsi élaboré un référentiel pratique destiné aux acteurs, afin de favoriser les questions d'éthique du numérique. Nous estimons que dans nombre de cas la notion de tiers de confiance doit être considérée comme un élément essentiel de la constitution de cette filière de santé. Nous souhaitons, par ce biais, encourager l'innovation et le développement de la médecine de prévention, qui s'appuie forcément sur l'utilisation de données de santé venant de différentes sources.

Voici, pour moi et pour les adhérents de l'organisation que je représente, les éléments essentiels de ce tournant numérique en cours, que nous saluons.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, Mme Zablit-Schmitz. Nous aurons l'occasion de revenir sur tous ces thèmes, qui sont au cœur du sujet d'aujourd'hui. Le regard de Syntec Numérique est précieux.

Je vais maintenant donner la parole à Mme Jeanne Bossi-Malafosse, avocate associée au sein du cabinet Delsol avocats. Nous avons, Madame, eu l'occasion de vous entendre dans le cadre des travaux relatifs à la loi sur les données personnelles et à la transcription du règlement général de protection des données. Vous êtes une experte des enjeux juridiques de la collecte de données. Se posent en effet des questions liées à la nature des données, au statut juridique de données pseudonymisées, aux modalités de collecte, au consentement, à la reconnaissance de dispense d'information reconnue par le RGPD, au référentiel de sécurité, à l'interopérabilité, à l'usage des algorithmes ou encore aux montages juridiques et légaux de toute sorte, autant d'aspects sur lesquels votre concours sera précieux.

**Mme Jeanne Bossi-Malafosse, avocate associée, Delsol avocats.-** Je souhaite, dans le temps relativement bref qui m'est imparti, aborder quatre points qui me semblent essentiels et constituent des sujets juridiques au sens large.

Le premier est celui du statut de la donnée pseudonymisée. M. Gissot a rappelé précédemment que, dans le système national des données de santé, les données concernées étaient pseudonymisées. Qualifiées auparavant de « données indirectement nominatives », elles ne permettent pas, loin s'en faut, l'identification directe de la personne. Or aujourd'hui, ces données pseudonymisées sont juridiquement considérées comme des données à caractère personnel, donc soumises aux règles d'application du règlement général sur la protection des données et de la loi Informatique et libertés modifiée. Il me semble qu'il serait utile de réfléchir davantage à ce statut particulier de la donnée pseudonymisée, dans la mesure où les appréciations restent à ce propos encore divergentes, en particulier lorsqu'il s'agit d'accéder à ces données : comment définit-on le bon niveau d'agrégation de ces données quand elles sont transmises à quelqu'un qui en respecte les conditions ? L'obligation posée d'autre part par le règlement européen de mener une analyse d'impact, c'est-à-dire de constater et d'assumer le risque résiduel d'identification, ne devrait-elle pas être davantage étudiée, pour ne pas se trouver avec des niveaux d'agrégation destructeurs de la valeur des recherches susceptibles d'être menées ? Ce point me paraît devoir être interrogé.

Concernant l'information et le consentement, fort bien présentés par David Gruson, rappelons que le consentement n'est pas, à l'heure actuelle, juridiquement exigé dans tous les cas, loin s'en faut. N'instaurons pas, par conséquent, un consentement général alors que les textes ne l'exigent pas aujourd'hui dans tous les cas. En matière d'information, puisque nous devons appliquer le règlement général d'avril 2016, rappelons que celui-ci prévoit deux situations lorsque l'on entend réutiliser des données : soit la personne a été informée au départ que ses données pouvaient être réutilisées, soit on se dispense de cette nouvelle information. Bien que la CNIL ait mené un important travail de simplification, on constate aujourd'hui dans certains de ses textes, dont une méthodologie qu'elle a fort utilement adoptée l'été dernier sur les recherches

impliquant la personne humaine et concernant la réutilisation de données ou d'échantillons biologiques, que, dès lors que l'on veut déroger à l'obligation individuelle d'information, on est obligé de déposer à nouveau une demande d'autorisation auprès de la CNIL, ce qui ne facilite pas l'accès à ces données. Peut-être faudrait-il revoir ce point afin d'alléger les formalités, sous réserve évidemment que toutes les garanties appropriées soient prises.

Le troisième sujet concerne l'incitation au partage des données et les référentiels de sécurité et d'interopérabilité. Nous avons constaté dans notre pays, depuis plusieurs années, que le secteur de la santé n'était pas propice au partage et à l'échange des données. Généralement, les structures détentrices de données entendent les garder et ne les partagent pas si elles n'y sont pas obligées. À partir du moment où l'on souhaite inclure dans le SNDS l'ensemble des données cliniques pour pouvoir enfin mener des études en vie réelle, il va falloir des incitations fortes pour que les professionnels de santé partagent leurs données.

Ceci a également un rapport avec la gouvernance du *Health data hub*, dont il a été question en début de matinée et sur lequel nous reviendrons certainement plus longuement lors de la seconde table ronde. Il est notamment important de bien réfléchir au statut juridique de cette plateforme et de savoir par exemple si le statut de GIP est le mieux adapté.

Je voudrais particulièrement insister sur les référentiels de sécurité et d'interopérabilité. Il n'y aura pas de SNDS avec les données en vie réelle, ni d'Espace numérique de santé tel que prévu par le projet de loi, si l'on ne rend pas opposables des référentiels de sécurité et d'interopérabilité clairs et lisibles par les acteurs. De tels référentiels figuraient dans la loi HPST de 2009 et avaient été confirmés dans la loi de 2016. Pour autant, on ne dispose toujours pas des arrêtés d'opposabilité. Je puis vous affirmer, pour étudier des cas concrets, que cette situation sert assurément certains acteurs. En effet, si certains respectent ces référentiels, avec des systèmes d'identification et d'authentification forts, d'autres s'en dispensent dans la mesure où les arrêtés d'opposabilité n'ont pas été publiés. Or cet élément est pourtant très important, surtout à l'heure où le règlement européen pose un droit à la portabilité des données, permettant à chacun d'exiger que l'on transfère ses données à un autre responsable de traitement ou que celles-ci lui soient restituées. Ce sujet de l'opposabilité des référentiels de sécurité et d'interopérabilité est absolument fondamental.

Le quatrième sujet que je souhaite aborder est celui des algorithmes. Il faut savoir que figure dans la loi Informatique et libertés, en particulier dans l'alinéa 2 de son article premier, la reconnaissance pour la personne de la possibilité de rester toujours maîtresse de ses données. Ceci rejoint l'autodétermination informationnelle mise en avant par les Allemands. Cette loi comporte également, dans son article 10, l'interdiction de prendre toute décision à l'égard d'une personne sans qu'elle en soit informée, disposition qu'il convient aujourd'hui d'apprécier au regard de celle du RGPD sur le profilage, qui rappelle la

nécessité d'une intervention humaine. Ces éléments sont essentiels dans les projets menés aujourd'hui. Peut-être pourrait-on les inclure dans les référentiels éthiques visés par le projet de loi. Il faudra en outre veiller, dans le projet de loi de transformation de notre système de santé et dans le cadre de la révision de la loi relative à la bioéthique, à ce qu'il n'y ait pas une multiplication des référentiels, notamment dans ce domaine qui est une zone grise et un peu molle, bien que tout à fait passionnante. Quelle sera l'autorité en charge de définir ces référentiels ? Quelle sera la légitimité de cette instance ?

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci, Mme Bossi-Malafosse. Pourrions-nous, dans la mesure où ces arrêtés d'opposabilité des référentiels dépendent de son ministère, avoir une réaction de M. Aubert ?

**M. Jean-Marc Aubert.-** Il faudrait aborder ce sujet, qui concerne des questions de droit et de protection du patient, avec le ministère de la justice. Il est impératif que nous fassions évoluer notre droit dans ce domaine, pour permettre l'usage des données dans les meilleures conditions. Les associations de patients et les académies avaient, me semble-t-il, beaucoup travaillé sur le RGPD pour obtenir une balance entre la capacité à utiliser ces données et la protection des patients. Certains amendements au RGPD déposés au Parlement européen proposaient de favoriser la protection des patients, au détriment des usages des données. Or les patients avaient été en première ligne pour soutenir les instituts académiques en vue de revenir à un équilibre.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Ces échanges montrent combien le sujet est important et complexe. Sans doute aurons-nous l'occasion d'y revenir.

Vous évoquiez, Mme Bossi-Malafosse, la question du statut juridique de GIP choisi pour la plateforme *Health data hub*. J'ai vu que ceci faisait partie des questions posées par les internautes et recueillies par ma collègue Valeria Faure-Muntian. Je vous propose d'y revenir tout à l'heure. Nous avons entendu, de la part des uns et des autres, des appels à la souplesse : il importe dans cette logique de se demander si la structure telle qu'elle est envisagée sera bien adaptée à une telle ouverture.

Je vais sans plus tarder donner la parole à M. Thomas Dautieu, directeur de la conformité, et M. Erik Boucher de Crèvecoeur, ingénieur expert à la direction des technologies et de l'innovation de la CNIL. Il était bien évidemment naturel et indispensable de convier à cette table ronde les représentants du régulateur très respecté qu'est la Commission nationale de l'informatique et des libertés. Vous êtes accompagnés de Mme Tiphaine Havel, conseillère pour les questions institutionnelles et parlementaires à la présidence de la CNIL. Nous vous écoutons sur les questions et enjeux du développement des algorithmes et de l'intelligence

artificielle en matière de protection des données, notamment des données de santé, aspects éthiques inclus.

**M. Thomas Dautieu, directeur de la conformité, CNIL.-** Merci d'avoir convié la CNIL, autorité administrative indépendante en charge de la protection des données, à cette table ronde.

Nous constatons, depuis une dizaine d'années, un double mouvement, avec, d'une part, l'émergence et la multiplication des données massives, du *big data*, et, d'autre part, le développement de moyens techniques permettant une exploitation de plus en plus puissante de ces données. Nous sommes ainsi à un moment charnière de bouleversement du traitement des données, y compris dans le domaine de la santé. Toutes ces évolutions offrent évidemment des perspectives formidables en termes de médecine prédictive, de diagnostic, mais posent aussi un certain nombre de questions en matière de protection des données, sur lesquelles je vais centrer mon propos.

Les algorithmes, l'intelligence artificielle, ont besoin de beaucoup de données pour apprendre, fonctionner, s'améliorer. Or le plus souvent, les algorithmes préfèrent pour ce faire disposer de données non anonymes, c'est-à-dire soit directement personnelles, soit pseudonymisées, afin de pouvoir les chaîner entre elles et établir des corrélations et des liens. À partir de là, on entre, face à ces données pseudonymisées ou personnelles, dans un cadre protecteur, prévu, d'une part, par le règlement général européen sur la protection des données, d'autre part, par la loi Informatique et libertés, dans la mesure où le législateur français a fait le choix, en juin dernier, de prévoir des dispositions spécifiques en matière de traitement des données de santé. C'est dans ce cadre protecteur que l'on doit traiter les données de santé, juridiquement qualifiées de « sensibles ». Qu'y a-t-il en effet de plus parlant qu'une donnée de santé par rapport à la vie privée des personnes concernées ?

En matière de SNDS, les prémices de ce cadre exigeant, qui va évoluer avec le projet de loi déposé voici quelques jours, ont été posés par la loi dite « Touraine » de janvier 2016.

La CNIL identifie dans ce contexte un certain nombre de problématiques, dont deux nous paraissent essentielles. La première est l'information des personnes. Le traitement des données de santé ne pourra s'effectuer sereinement que dans un cadre de confiance partagée entre les patients, les professionnels de santé et les industriels. Pour que cette confiance s'installe et que les patients acceptent que leurs données soient traitées, il faut leur délivrer une information quant au devenir et au traitement de ces données, afin qu'ils puissent dans un second temps, le cas échéant, exercer leurs droits : droit d'accès, de rectification, d'opposition, que nous connaissons en France depuis plus de quarante ans, mais aussi désormais droit de portabilité. Cette question de l'information des personnes est, de notre point de vue, tout à fait centrale. Elle renvoie à la notion de

transparence portée par le règlement général sur la protection des données. Ceci est d'autant plus important que l'on assiste actuellement à l'émergence d'acteurs privés, dont le modèle économique repose sur la revente de données de santé, alors que ce secteur était jusqu'à présent essentiellement le fait d'opérateurs publics. Certains fournissent des objets connectés, des applications, d'autres des tests génétiques ; mais le *business model* réel en arrière-plan repose toujours sur la revente des données de santé collectées. Nous n'avons pas à porter d'appréciation sur cette mécanique, mais il nous importe que les personnes à l'origine des données sachent ce qu'il va en advenir.

J'attire votre attention sur le fait qu'information ne signifie pas consentement. L'AP-HP a par exemple souhaité constituer un entrepôt de données pour mener des études ultérieurement. Plus de huit millions de personnes ont ainsi vu leurs données versées dans cet entrepôt. Juridiquement, les textes n'exigent pas un consentement de ces personnes. Ces dernières ont simplement dû être informées, afin de pouvoir, le cas échéant, s'opposer au transfert de leurs données dans cet entrepôt.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Quel est, en pratique, le taux de rejet ?

**M. Thomas Dautieu.-** Je n'utiliserais pas le terme de « rejet ». Les personnes ont été informées individuellement et très peu d'entre elles se sont opposées. Je ne dispose pas des chiffres précis, mais puis vous dire que cette situation est marginale. J'y vois le signe que, dans une démarche de confiance, les patients perçoivent l'intérêt que peut présenter l'utilisation de leurs données à des fins de recherche.

Le deuxième élément essentiel à nos yeux en termes de protection des données, au-delà de l'information, concerne la sécurité des données. Les bases de données de santé, sensibles par nature, se multiplient et la future loi va permettre des appariements entre des entrepôts, des bases de données et les données issues du SNDS. L'AP-HP pourrait par exemple demander à apparier son entrepôt de données avec des données issues du SNDS. Il est essentiel que la sécurité des données soit totalement assurée, dans une optique de confiance. Rien ne serait plus redoutable pour cette logique d'utilisation des données de santé qu'une fuite de données de santé massive, rendues ainsi accessibles à des tiers non autorisés. La sécurité informatique est indispensable. Elle a un coût, prend du temps, mais est essentielle pour créer un climat de confiance entre patients, industriels et professionnels de santé. Il faut selon nous trouver un point d'équilibre entre, d'une part, la nécessaire circulation et l'utilisation des données à caractère personnel, d'autre part, la protection de ces données. Le rôle de la CNIL en la matière est évidemment d'accompagner l'ensemble des acteurs de cet écosystème, publics comme privés, grâce notamment à des textes simplifiant les procédures prévues par le législateur.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, M. Dautieu. Nous avons pu voir, au fil de nos travaux sur ces questions, à quel point la CNIL évoluait en prenant en compte la nécessité de trouver des solutions efficaces et pragmatiques. Votre exposé en témoigne et nous en sommes très heureux.

Vous avez insisté sur l'importance de la confiance partagée entre les patients, les professionnels de santé et les industriels. Ceci constitue une transition idéale pour donner la parole aux patients, par l'intermédiaire de M. Gérard Raymond, vice-président de la fédération d'associations de patients France Assos Santé. Notre sujet impose de placer l'humain au cœur du système et de la réflexion algorithmique. D'une façon générale, il est important, dans les questions relatives à la santé, d'entendre la voix des patients, qui sont assurément les premiers concernés. Je rappelle que France Assos Santé a vu sa mission officiellement reconnue par son inscription dans le code de la santé publique, par une loi de janvier 2016. Cette fédération a été créée par la réunion de 72 associations nationales fondatrices. Elle compte aujourd'hui 80 associations membres et constitue donc un représentant incontournable et fédérateur d'associations de patients et d'utilisateurs du système de santé.

France Assos Santé a publié un communiqué de presse après le dépôt du projet de loi relatif à l'organisation du système de santé, évoquant en particulier les questions de télésoins et de plateformes d'accès aux données de santé. Nous serions très intéressés de connaître votre point de vue sur ces différents sujets.

**M. Gérard Raymond, vice-président de la fédération d'associations de patients France Assos Santé.-** Je tiens en préambule à préciser que je ne suis qu'un profane et ne peux en aucune manière rivaliser avec les expertises présentes autour de cette table.

Mes propos porteront sur deux axes. Le premier, évoqué notamment par le représentant de la CNAM, concerne le fait que nous disposons de beaucoup de données cliniques, de données d'établissements. Il manque toutefois, parmi les données que nous pouvons recueillir aujourd'hui, un pan important qui est celui des données de la vie réelle des patients, que d'aucuns qualifient d'« expérience patients ». Le deuxième axe sera consacré à la co-construction, à la confiance à établir entre l'ensemble des partenaires, afin de pouvoir réussir cette transformation numérique qui va impacter et impacte déjà notre société.

Pour ce qui est du recueil des données, on ne peut se contenter aujourd'hui, si l'on veut mener des travaux de recherche, de ne disposer que de données cliniques. Il faut aussi que nous ayons des données en vie réelle des patients, des informations sur la façon dont chacun de nous perçoit sa qualité de vie, évalue et apprécie le système de santé au regard des propositions qui peuvent lui être faites. Nous considérons que la France est, de ce point de vue, largement en retard par rapport à nombre d'autres pays. Si nous voulons combler cet écart, il faut

insister lourdement sur le recueil de ces données en vie réelle. Certaines des associations de patients que je représente se sont déjà dotées de moyens en ce sens, en créant leurs propres plateformes pour recueillir les attentes et les besoins des patients. Il n'empêche que nous avons aujourd'hui à travailler sur les questionnaires, que ce soit avec la Haute autorité de santé ou d'autres services au niveau du ministère. Il nous faut réfléchir ensemble à la manière d'améliorer le recueil des données en vie réelle et l'évaluation de notre système de santé issue de l'expérience patients. Soyez persuadés que les associations de patients que je représente souhaitent évidemment être parties prenantes de cet engagement, qui va dans le sens de l'évolution de notre système de santé.

Le deuxième axe, qui rejoint le premier, concerne la démarche de co-construction. Il a été question précédemment de sécurisation des données, de consentement, de confiance. Pour nous, le débat se situe à deux niveaux. Individuellement, chacun d'entre nous, citoyen, patient, peut donner son consentement ou voir parfois ses données être utilisées sans son accord. Ces données lui appartiennent, mais il les laisse à disposition, avec ou sans son consentement. Il est pour nous extrêmement important que la personne sache qu'elle est représentée par des structures reconnues aujourd'hui d'utilité publique par la loi, agréées par le système de santé. Il faut systématiquement associer les associations de patients à la co-construction de l'ensemble des dispositifs, qu'il s'agisse des plateformes dans le cadre du service public, avec aujourd'hui une grande simplification, ou d'autres plateformes de recueil des données. Il est indispensable que nous soyons partie prenante, pour assurer la confiance à laquelle faisait référence à juste titre le représentant de la CNIL, grâce à l'information légitime que doivent recevoir l'ensemble des patients. Cette prise de responsabilité des associations de patients est pour nous un acte fort de la réussite, aujourd'hui et demain, de l'utilisation de l'ensemble de ces données par l'intelligence artificielle.

Une association que je connais bien a publié, voici quelques mois, quinze propositions pour une médecine fondée sur l'humanisme. Plusieurs de ces recommandations portent sur les nouvelles technologies. Nous sommes persuadés que la collecte des données de santé et demain leur utilisation pour bénéficier d'un vrai registre de l'état de santé de nos populations et bâtir un système de santé fondé sur les attentes, les besoins, les services attendus par nos concitoyens et non simplement sur une offre, sont essentielles à l'évolution de notre système de santé. Il nous apparaît indispensable de travailler à rapprocher les personnes, soignants et soignés, soignants entre eux, professionnels de santé, régulateurs et payeurs. Tous les outils mis en œuvre doivent converger vers cette nécessité de coopération, de coordination et de travail en commun.

Le sens que nous devons donner à l'ensemble des dispositifs, notamment d'intelligence artificielle, qu'il nous faut élaborer aujourd'hui pour le recueil des données de santé et leur utilisation est de contribuer au rapprochement des hommes



et des femmes acteurs de ce système, afin que se crée entre tous une véritable confiance.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, M. Raymond.

### **Débat et questions du public**

**M. Cédric Villani, député, premier vice-président de l'Office.-** Après cette succession de prises de parole, le moment du débat est venu. Sont présents ici des membres de l'Office parlementaire d'évaluation des choix scientifiques et technologiques, mais aussi des membres de la commission chargée des affaires sociales, qui examinera bientôt le projet de loi que nous nous apprêtons à évoquer. Tous auront certainement à cœur de participer à ce débat. Je leur laisse donc la parole, afin qu'ils nous fassent part de leurs questions et remarques. Les invités de la seconde table ronde sont bien évidemment également invités à s'exprimer au cours de cette discussion, chaque fois qu'ils l'estimeront utile ou nécessaire. Je vous rappelle par ailleurs que ma collègue Valéria Faure-Muntian nous présentera une synthèse des questions soumises par les internautes, après les présentations liminaires des intervenants de cette première table ronde.

**Mme Albane Gaillot, députée.-** Je suis commissaire aux affaires sociales et responsable pour le groupe La République en Marche des titres 2 et 3 du projet de loi santé. Je trouve nos échanges très intéressants et éclairants dans le cadre de l'étude de ce projet de loi. Les données de santé constituent un enjeu important. La création du *Health data hub*, permettant une collecte organisée, pilotée des données apparaît aujourd'hui comme une nécessité. Ne faudrait-il pas que ce *hub* soit un outil au service de la filière santé, un coordinateur, un animateur de la filière et pas seulement un outil technique ?

J'ai bien entendu par ailleurs toutes les remarques sur la co-construction, émanant à la fois des associations de patients et des acteurs privés par la voix de la représentante de Syntec Numérique. Il me semble important de réfléchir au mode de gouvernance et à la manière de construire la démarche et le système.

Au niveau technique, ce *Health data hub* pourra-t-il être assez souple et interopérable entre les acteurs ?

**M. Cédric Villani, député, premier vice-président de l'Office.-** Nous avons entendu parler du *Health data hub* comme devant être un groupement d'intérêt public, dans la suite de l'INDS. Mme Bossi-Malafosse a posé la question de savoir s'il s'agissait là du mode de gouvernance le mieux adapté au regard de la souplesse exigée et du rôle d'animation que vous évoquez. On se souvient qu'il a fallu plus d'un an pour mettre en place l'INDS dans le cadre assez contraignant du

GIP, qui requiert un accord *ex ante* de tous les acteurs. La structure telle qu'elle est envisagée ne risque-t-elle pas d'être un frein à l'efficacité de cette plateforme ?

**Mme Jeanne Bossi-Malafosse.-** Ayant suivi ces questions depuis plusieurs années et à différents titres, j'ai pu observer une tendance consistant à créer des groupements d'intérêt public pour gérer des sujets particuliers : on a ainsi créé le GIP CPS dans le cadre de la mise en place de la carte de professionnel de santé, le GIP DMP en 2004 lorsqu'il s'est agi de lancer le dossier médical personnel, le GMSIH sous forme de GIP pour la modernisation des systèmes d'information hospitaliers. Une première politique d'harmonisation a été menée pour aller vers davantage d'efficacité, avec la création, également sous forme de GIP, de l'Agence des systèmes d'information partagés en santé (ASIP Santé), dont l'objet était justement de fusionner tous ces GIP. On avait toutefois pris soin, lors de la création de cette instance dans la loi HPST, de préciser, comme dans le projet de loi actuel à propos du *Health data hub*, qu'il était possible de recruter des personnes dans le secteur privé, afin de bénéficier de leurs compétences. L'étape suivante consistant en la mise en place d'une grande agence de santé publique gérant tous ces sujets n'a jamais été franchie, pour des raisons sur lesquelles nous ne reviendrons pas ici. La création d'un GIP demande du temps et est assez compliquée. Il est en outre très important de clairement en définir la gouvernance. Si l'on intègre les données de santé produites par les médecins dans le cadre des dossiers médicaux afin de mener des études en vie réelle, comment cette gouvernance va-t-elle s'organiser ? Va-t-on continuer à avoir une gouvernance très étatique, qui ne participe pas toujours à l'efficacité de la mise à disposition des données de santé ? Ne serait-il pas préférable dans certains cas d'envisager une formule juridique davantage ouverte vers la société, permettant de gérer des activités concurrentielles, qui doivent aller vers l'exploitation de données ? Le fait de ne pas accéder aux données de santé, de ne pas partager, de ne pas échanger est, comme l'a rappelé David Gruson, un facteur négatif pour la santé.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Quelles alternatives avez-vous en tête, qui pourraient répondre aux objectifs et critères que vous mentionnez ?

**Mme Jeanne Bossi-Malafosse.-** En termes de gouvernance, le GIP est une solution ; la société par actions simplifiée (SAS) en est une autre. Peut-être existe-t-il d'autres possibilités. L'idée de maintenir les fonctions régaliennes fondamentales visant à assurer aux patients, aux citoyens, la sécurisation de leurs données et le respect de la protection des données personnelles est acquise et ne fait pas débat. Peut-être faudrait-il permettre, par la création d'une entité dépendant éventuellement du GIP sur ces principes régaliens, d'aller davantage vers des activités concurrentielles et l'exploitation de données, en maintenant l'État comme « actionnaire majoritaire ». Ceci offrirait une plus grande souplesse de fonctionnement.

**Mme Angèle Prévile, sénatrice.-** Merci à toutes et tous d'être présents et de nous éclairer sur ces sujets complexes. M. Gissot a évoqué les « données patients » et M. Raymond la « vie réelle des patients » : s'agit-il de la même chose ?

Tous ces dispositifs ont pour objectif d'améliorer les soins dispensés aux patients. Ne pourrait-on par conséquent intégrer dès le départ dans le système le retour d'expérience des patients sur leur vécu vis-à-vis des soins qui leur ont été prodigués ? Comment, dans le cas contraire, mesurer l'amélioration du système de santé ?

**M. Cédric Villani, député, premier vice-président de l'Office.-** Je vous propose de collecter plusieurs questions, puis d'écouter les réponses des uns et des autres.

**M. Bruno Sido, sénateur.-** Ces données doivent servir à établir des statistiques, à construire des algorithmes permettant d'être plus précis et rapides dans les diagnostics par exemple. Pourquoi a-t-on par conséquent besoin de connaître le nom du patient ou de pseudonymiser les données ? Pourquoi ne pas simplement les anonymiser ?

**M. Thomas Gassilloud, député.-** Merci pour ces interventions très éclairantes. Je ne suis absolument pas spécialiste des sujets de santé et travaille plutôt sur les questions de défense nationale, notamment sur l'armée de terre. J'ai été étonné de constater, en vous écoutant, à quel point les réflexions en matière de défense et de santé étaient proches. Ceci n'est finalement pas si surprenant, dans la mesure où, dans les deux cas, la vie est souvent en jeu : on a donc besoin d'une éthique forte, de prendre parfois des décisions rapidement. La crainte quant au partage des données et les questions relatives à l'articulation avec les industriels, à l'interopérabilité des systèmes, sont également du même ordre.

Je m'interroge pour ma part sur notre capacité collective à faire en sorte que notre puissance publique, en l'occurrence notre système de santé collectif, puisse rester dans la course par rapport aux acteurs privés. L'un des premiers risques évoqués est celui de la sous-efficience liée au fait de ne pas utiliser l'IA en matière de santé. À plus long terme, au-delà de la sous-efficience, c'est même notre système de solidarité collective qui pourrait être remis en cause. La CNIL a fait part de l'émergence d'acteurs privés dont le modèle économique repose sur la revente des données de santé. Il me semble que l'écosystème qui se met en place, avec le développement des objets connectés, est de nature à favoriser le développement de *business models* faisant entrer massivement de nouveaux acteurs dans ces domaines. De la même manière, on a vu ces dernières années, dans le monde de la défense, des organisations non étatiques développer des capacités qui mettent en grande difficulté les puissances étatiques occidentales, grâce à l'appropriation massive de technologies. N'y a-t-il pas un risque lié au fait que la technologie puisse, à terme, remettre en cause la pérennité de nos systèmes de solidarité

collective, en raison de l'entrée massive d'acteurs privés ? Comment, collectivement, s'en prémunir ?

**M. Cédric Villani, député, premier vice-président de l'Office.-** Je souhaiterais pour ma part rappeler que l'utilisation de modèles de sociétés par actions simplifiées dans le secteur public s'est généralisée ces dernières années dans certains domaines. Ainsi, la French Tech, à Station F., est une SAS, tout comme certaines structures d'accélération de transfert technologique et la plupart des instituts pour la transition énergétique. On pense également à API-AGRO, à France Brevets, bientôt à TeraLab, ou encore à des fondations de coopération scientifique dotées d'une filiale sous forme de SAS. Il serait intéressant que certaines personnes ici, plus expertes que moi en gouvernance et en montage juridico-légal, puissent commenter cet élément.

**M. David Gruson.-** Je souhaiterais, en écho aux propos de Mme Gaillot et de M. Gassilloud, souligner l'importance de considérer le facteur temps. Sans doute faudra-t-il, dans le débat parlementaire que vous mènerez, veiller, au niveau national et européen, à ne pas sur-complexifier le dispositif. L'idée que la création du *Health data hub* serait un objectif d'intérêt économique ou industriel va de soi. Cet outil, sous forme de plateforme d'intérêt public, représente évidemment une avancée éthique, mais est aussi un vecteur de soutien à l'émergence de solutions de médecine algorithmique nationales. Ceci nécessite toutefois d'être porté politiquement. Il s'agit moins, selon moi, d'une question de droit que de volonté politique.

Il m'apparaît, pour répondre à M. Gassilloud, que le risque pour le système de solidarité et notre sécurité sociale tient moins à l'émergence d'acteurs privés qu'à la capacité des professionnels et des patients d'éviter, par le numérique, notre système de santé et le principe de solidarité. L'exemple du règlement général sur la protection des données témoigne d'une initiative européenne communautaire forte. Des acteurs privés, y compris de très gros « calibre » mondial, se sont soumis à ce nouveau cadre normatif. Ceci n'est donc pas impossible dans l'absolu, y compris sur le terrain de la régulation positive. En revanche, si nous tardions trop à avancer sur le partage des données dans un cadre d'intérêt public national, idéalement européen (le *Health data hub* étant, espérons-le, une première étape dans cette direction), les conséquences seraient très négatives.

Ainsi, l'agence de sécurité sanitaire américaine, la FDA, a autorisé en avril dernier l'utilisation d'un algorithme de reconnaissance d'images pour les rétinopathies diabétiques. On peut ainsi très facilement imaginer le cas d'usage d'un patient français qui, plutôt que de patienter plusieurs mois avant d'obtenir un rendez-vous avec un ophtalmologue, aura recours à cet algorithme, en envoyant ses données de l'autre côté de l'Atlantique, moyennant paiement. Ceci met en lumière un risque de rupture de la protection des données à l'initiative des patients et des professionnels.

Il ne faut pas perdre de vue que les risques pour la protection des données sont aussi alimentés par chacun de nous ; il ne s'agit pas seulement d'un sujet de droit, mais d'une question de comportement, donc de formation et de débat public. Il apparaît également que ceux parmi nos concitoyens qui pourront se permettre ce genre de comportement seront ceux qui en auront les moyens, avec de façon sous-jacente un risque de délitement des régimes de sécurité sociale qui sont à la base de notre système. Mon propos peut sembler théorique, mais il s'agit en réalité déjà d'un sujet pratique. C'est la raison pour laquelle il est vraiment essentiel que le cadre juridique français et européen s'adapte au principe de partage des données, avec les éléments de protection juridique qui ont été rappelés par divers orateurs. Il est nécessaire d'agir suffisamment vite et lisiblement pour ne pas accélérer en pratique ces comportements d'évitement du système.

**M. Cédric Villani, député, premier vice-président de l'Office.-** En clair, si nous sommes trop protecteurs, pas assez rapides et trop lourds, nous aboutirons à l'effet inverse de celui recherché, car les solutions sauvages, privées, porteuses d'inégalités, se multiplieront, en évitement du système.

J'ai vu que M. Olivier Clatz, intervenant de la seconde table ronde, avait demandé la parole. Fondateur de Therapixel, M. Clatz dispose d'une solide expérience en matière de développement de la technologie au service de la santé, puisque sa société s'est fortement illustrée au niveau international dans ce domaine.

**M. Olivier Clatz, fondateur de Therapixel, start-up d'édition de logiciels d'imagerie médicale.-** L'une des questions portait sur les notions d'anonymisation et de pseudonymisation. Je ne connais pas aujourd'hui de recherche ou de société ayant vraiment besoin d'accéder au nom des patients. En tant qu'industriel, je n'ai jamais eu affaire à une telle demande. Le terme de « pseudonymisation » est utilisé lorsqu'il est possible, par un moyen quelconque, de revenir à l'identité de la personne, ce qui est souvent le cas sur des données volumineuses, par corrélation.

Le questionnement s'est aussi largement orienté vers le statut de GIP. Je ne suis pas expert en droit, mais puis vous dire qu'accéder à ces données, les traiter, les exploiter requiert le recours à des professionnels disposant de fortes compétences. Nous avons besoin de recruter des talents, d'acheter des machines et des matériels spécifiques et probablement pas classiques dans l'administration. L'une des vocations du *Health data hub* est aussi de travailler avec des entreprises comme la nôtre. La structure doit le permettre et l'assumer. Ceci suppose un choix politique, ainsi que l'a souligné David Gruson. Il est très important d'assumer cette volonté de faire participer l'écosystème des entreprises à cette démarche. Le modèle des IRT mentionné par Cédric Villani est, de ce point de vue, très intéressant, avec des fondations de coopération scientifique qui se différencient en une partie consacrée à l'intérêt commun et une partie sous forme de société, filiale à 100 % des IRT, qui leur permet de développer leurs intérêts économiques. Il s'agit probablement d'un modèle à déployer.

**M. Jean-Marc Aubert.-** Nous avons bien conscience du fait que l'animation sera au cœur du travail du *Health data hub*, ce qui pose d'ailleurs la question de la coopération avec l'ensemble de l'écosystème. L'intérêt du GIP est de permettre, au moment de la convention constitutive, de réunir plusieurs partenaires ; celui d'une SAS d'offrir davantage de souplesse sur la possibilité d'intégrer un partenaire nouveau, puisqu'il suffit pour cela d'accorder quelques droits de vote ou une petite partie du capital. Le point important dans le choix, qui a motivé la prise de position de la ministre en faveur du GIP et dont il appartiendra à l'Assemblée de débattre, a été de privilégier la garantie, pour nos concitoyens, que le statut public reste majoritaire dans cette structure. On parle en effet de données de santé, qui sont des données publiques, et l'on sait que d'aucuns s'inquiètent de l'intrusion potentielle de grands groupes n'agissant pas toujours avec l'éthique souhaitable.

Les données du *Health data hub* seront en outre des données anonymisées et aucun usage ne cherchera à les réindividualiser. On restera sur des usages anonymes. Seuls certains algorithmes ou applications développés sur ces données pourront les utiliser de façon nominative, mais toujours dans le cadre d'un consentement explicite du patient et en les sortant dans ce cas du *hub*.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Je comprends parfaitement l'importance de garantir le caractère public de la structure, quelle qu'elle soit. Ceci ne pourrait-il pas être imposé dans la définition d'une structure de type SAS ou FCS-SAS, comme évoqué par Olivier Clatz ? Il s'agirait alors d'inscrire dans la loi que tel pourcentage (majoritaire, très majoritaire ou totalité) du capital ou de la gouvernance doit être détenu par la puissance publique.

**M. Gérard Raymond.-** Il faudrait que l'État soit pour le moins largement majoritaire, même si la structure a vocation à être largement ouverte et la gouvernance à faire participer l'ensemble des acteurs. L'intérêt public nous semble évident, afin d'éviter que notre système de santé et notre médecine ne dérivent vers un système libéral à l'américaine.

**M. Gérard Longuet, sénateur, président de l'Office.-** Je suis très perplexe. Je ne suis pas spécialiste de ces questions, mais connais en revanche assez bien l'organisation de l'État. L'État bénéficie, sous le contrôle du législateur, du monopole de la contrainte. C'est là en réalité sa seule légitimité que de pouvoir imposer à des individus, des citoyens, des comportements fondés sur la base d'une loi, cette dernière étant définie par le Parlement, dont c'est la fonction. J'évoque à dessein cet élément dans la mesure où cette contrainte ne peut s'exercer que si elle est légitime et possible. Or nous sommes ici, à l'Office parlementaire, confrontés (comme l'ensemble de nos compatriotes, bien que ces derniers n'en aient peut-être pas autant conscience) au fait que, d'évidence, l'autorité de contrainte de l'État disparaît progressivement, et ce d'autant plus fortement dans une société mondialisée, numérique.

Vous indiquiez schématiquement, M. Aubert, que « qui paie commande ». Je n'entrerai pas dans ce débat, car en réalité ce n'est pas l'État, mais le contribuable qui paie. Cette légitimité n'est pas suffisante. En effet, en économie, la frontière se déplace en permanence et la démonstration effectuée à l'instant par M. Clatz montre que la société ne connaît pas, dans ce domaine, de frontières, ni les contraintes de l'État. Prenons l'exemple des tests ADN, encadrés en France par une loi et des règles claires, qui s'imposent à nous : ce cadre est parfaitement contrebattu par la réalité d'une commercialisation de tests ADN dans des conditions qui font que le citoyen échappe à l'autorité de la loi. Traditionnellement, l'image du contrevenant à la loi était celle du contrebandier avec son sac de victuailles, évitant les gabelous. Cette vision est anachronique. Aujourd'hui, l'outil du contrevenant est le « clic » d'achat sur internet. Ceci rejoint d'ailleurs le problème de la fiscalisation des GAFAs pour des prestations réalisées dans d'autres domaines et montre la limite de l'action de l'État. Je pense que votre approche est indispensable pour les parlementaires que nous sommes. Ne rien faire serait priver nos compatriotes d'un droit, d'une aspiration légitime à la santé ; mais agir de façon trop contrainte entraînerait cette évasion par la société numérique.

Pour répondre à mon collègue et ami Bruno Sido, je pense en effet que nous n'avons pas, aujourd'hui, besoin de connaître les noms des personnes auxquelles correspondent les données. Il m'apparaît toutefois qu'un jour viendra nécessairement où quelqu'un se dira que le fait de connaître les noms permettrait de générer une valeur ajoutée, en développant une activité rendant service et dont la contrepartie serait un résultat économique. Le mécénat n'est pas le moteur principal de nos sociétés sur le plan mondial et les promoteurs agissent en effet généralement avec l'espérance du profit. Il n'y a pas de nécessité scientifique à connaître l'identité des personnes, mais une nécessité de croiser des fichiers, donc de savoir de qui l'on parle, surtout si l'on introduit, comme le suggère M. Raymond, les données de la vie réelle. Je connais un peu le cas des mineurs de charbon. Il existe des mines avec ou sans silicose. Par conséquent, connaître le contexte et les gens peut être utile. Vous dites « qui paie commande » ; or ce sont les assureurs qui paient et essaient de ce fait de connaître les populations à risque, ce qui implique de descendre vers la connaissance physique de l'individu. La question est donc ouverte aujourd'hui. Cette table ronde est indispensable pour en débattre et le débat parlementaire sera éclairant ; mais l'affaire n'est pas aisée.

L'État, depuis un siècle environ, a toujours cherché des organismes extérieurs, afin de pouvoir coopérer avec des partenaires. Nous avons connu par exemple la génération des offices, dans les années 1930, celle des commissariats, des délégations, des agences, puis des commissions nationales et des autorités indépendantes. La SAS est une formule en réalité assez ambiguë : en effet, dans une société capitaliste, l'important réside dans les capitaux. Or je ne vois pas là les choses très clairement. Je pense en revanche qu'il faut que l'État accepte d'avoir l'attitude la plus ouverte et la plus attentive. On compte des réussites ; la CNIL en est une, car elle s'est imposée comme un acteur de régulation du système public.

Peut-être d'ailleurs devrait-elle jouer un rôle plus important dans la régulation de ces données. Il est absolument certain que nous ne ferons pas un système national qui se pense hermétique et serait indifférent à la réalité internationale. Peut-être manque-t-il dans cette table ronde un éclairage sur ce qui se fait ailleurs, afin de savoir comment d'autres pays, en Europe ou ailleurs, envisagent de progresser dans cette direction. Il ne me semble pas réaliste de penser que nous réglerons ce problème en ne le considérant qu'au niveau franco-français et étato-étatique. Nous avons besoin des partenaires privés. Les patients le demanderont.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Dans le cadre de l'appel à projets relatif au *Health data hub*, est-il prévu d'avoir des propositions de plateformes technologiques ? Ceci fera-t-il partie des questions soumises à appel d'offres ?

Le public nous demande par ailleurs à qui appartiennent les données collectées dans le cadre des dossiers pharmaceutiques. Où en est l'état de l'art sur la question de la transmission et de l'utilisation de ces données de dossiers, en fonction ou pas du consentement des patients ?

**M. Claude de Ganay, député.-** Ma question s'inscrit dans le prolongement de l'intervention de mon collègue Thomas Gassilloud. Un excellent rapport de l'Office paru voici un an et demi posait déjà la question de l'équilibre entre la recherche et la législation, la réglementation.

Je m'interroge par ailleurs sur les problèmes d'éthique en lien avec ces données, à partir du cas des grandes plateformes monopolistiques de type GAFA. J'ai beaucoup apprécié ce tour de table, mais m'interroge toujours sur ce problème d'équilibre entre l'utilisation des données à des fins mercantiles et commerciales et les usages relevant de l'intérêt général. Nous savons que la médecine évolue et sera de plus en plus individualisée. Comment effectuer ce *distinguo* ? Doit-on légiférer sur le sujet ?

**Mme Jeanne Bossi-Malafosse.-** Dans le cas du SNDS, les données sont pseudonymisées. N'oublions pas toutefois que les données concernées par le projet de loi sur la création de l'Espace numérique de santé seront nominatives. Toutes ces données, leur diffusion et leur accès sont protégés par la CNIL depuis quarante ans. Le système fonctionne plutôt bien. J'entends aujourd'hui des débats mettant en scène le « méchant secteur privé » vs. le secteur public. Allons au-delà de cela. Ce n'est pas parce que l'on est une entreprise du secteur privé que l'on va faire des choses terribles sur et avec les données. Il existe des garants, au premier rang desquels la CNIL. Ont par ailleurs été mis en place une série de comités. De nouveaux vont bientôt voir le jour. Je pense qu'il faut dépasser ce débat. Certaines sociétés privées demandent aujourd'hui à accéder à des données de santé pour les vendre : pourquoi l'interdirait-on, dans la mesure où ces entreprises se conforment au cadre de protection des données personnelles et agissent dans le respect de l'éthique ? Au nom de quel principe ? Ces entités sont, comme toute autre,



soumises au respect de la protection des données. Elles vendent d'ailleurs souvent des données aux pouvoirs publics eux-mêmes, qui souhaitent disposer de données en vie réelle. Regardons la réalité de la situation aujourd'hui, qui ne devrait plus opposer autant le public et le privé. Le projet de loi s'inscrit dans cette optique, puisqu'il met fin à un certain nombre d'oppositions issues de la loi de 2016, ce dont on ne peut que se réjouir. Il importe d'avancer sur ce sujet et de ne pas revenir sur des débats qui n'ont pas lieu d'être.

**M. Claude de Ganay, député.-** Mon propos n'avait pas pour objet d'opposer public et privé. Je suis conscient que la France est l'un des pays qui bénéficient de la meilleure réglementation, même si celle-ci s'élabore largement au plan européen. Nous avons toutefois vu, encore récemment, certaines entreprises des GAFAs condamnées pour utilisation des données personnelles. Jusqu'où peut-on aller dans ce domaine ?

**M. Gérard Raymond.-** Construire des murs n'a aucune valeur et n'offre aucune sécurisation. Je pense qu'il importe plutôt aujourd'hui de mettre en place de véritables contre-pouvoirs, sous la forme d'une participation de tous les acteurs, avec une totale transparence quant au rôle de chacun, à ses modalités d'action et à ses desseins. J'ai parfaitement conscience de la difficulté de l'exercice. Ceci m'apparaît toutefois, dans la nouvelle société qui est en train de se construire, comme une solution possible permettant de meilleures relations entre l'ensemble des acteurs.

**M. Thomas Dautieu.-** Vous avez souligné et interrogé le rôle de la CNIL par rapport à l'accès à ces données. Bien évidemment, la CNIL s'inscrit dans le cadre souhaité par le législateur. Vous avez, M. le président, fait référence aux assureurs : les assureurs ont aujourd'hui l'interdiction légale d'accéder aux données du SNDS dans le but par exemple de moduler leurs primes d'assurance. Le législateur a prévu cette garantie, dans la mise en œuvre de laquelle la CNIL joue un rôle de verrou ou de facilitateur, puisqu'il lui appartient d'autoriser ou pas l'accès au SNDS, dans le cadre de demandes ponctuelles, en veillant à ce que les garanties prévues par la loi soient respectées, à ce que les personnes soient informées ou en accordant des dérogations à l'information.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Quelqu'un souhaite-t-il s'exprimer sur la question du dossier pharmaceutique ?

**Mme Jeanne Bossi-Malafosse.-** Contrairement au DMP, construit comme étant le dossier de la personne elle-même, le DP (dossier pharmaceutique) a été conçu pour être l'outil du professionnel de santé, en l'occurrence du pharmacien, auquel il appartient d'en proposer l'ouverture au patient et de recueillir son consentement. Il est vrai qu'aujourd'hui le conseil national de l'ordre des pharmaciens, qui est responsable de ce dossier, réfléchit à son ouverture aux patients eux-mêmes, ce qui irait tout à fait dans la logique de la création prochaine de l'Espace numérique de santé, qui va permettre au patient d'accéder à ses

données, parmi lesquelles celles du dossier pharmaceutique, et de décider de les rendre ou non accessibles.

**M. Claude Gissot.-** Les données d'expérience patients recouvrent une grande variété d'éléments. Il peut s'agir par exemple d'un ressenti : est-ce que je me sens mieux après une opération qu'avant ? Il peut également s'agir du report, par le patient, *via* des applications dédiées, de mesures qu'il a lui-même effectuées, comme des mesures de tension artérielle par exemple. Ces données d'expérience patients sont des éléments d'évaluation, qui viendront compléter les données dont nous pourrions disposer par ailleurs.

Ceci renvoie à la question de la pseudonymisation et de l'anonymisation des données. Il est évident que les actions menées dans le cadre du futur *Health data hub* ou du SNDS actuel s'effectuent en mode pseudonymisé, car l'objectif est d'obtenir des statistiques, des indicateurs globaux et non des informations visant à changer le parcours individualisé, nominatif, du patient. Néanmoins tous les algorithmes que nous serons en capacité d'utiliser dans le SNDS pourront, à un moment donné, s'appliquer à un individu, dans le cadre de l'équipe de soin amenée à le prendre en charge. Il s'agit de deux schémas distincts, l'un visant à acquérir des connaissances d'ordre statistique, l'autre à appliquer éventuellement ce que l'on a appris à des cas individuels. L'idée première, qu'il s'agisse du SNDS ou du *Health data hub*, a toujours été de travailler en mode pseudonymisé, pour apprendre sur le système. L'application ultérieure est un autre débat.

**M. Bernard Nordlinger, chirurgien oncologue, Hôpital Ambroise Paré, Académie nationale de médecine.-** Je souhaitais intervenir dans le même sens que Mme Bossi-Malafosse : je pense qu'il faut faire preuve de subtilité et ne pas opposer public et privé. Certes il convient d'éviter que les données publiques soient utilisées par exemple par des assureurs pour adapter leurs cotisations en fonction des risques ; mais il ne faut pas oublier que l'essentiel des progrès médicaux qui font que l'espérance de vie n'est plus de 35 ans, mais a plus que doublé, sont dus à un cumul de progrès effectués par des équipes de chercheurs publiques et privées, et que l'industrie du médicament, si elle a pour but, comme toute société, de faire des bénéfices, apporte aussi sur le marché des médicaments qui permettent des progrès considérables. Il faut s'adapter à chaque cas et ne pas verser systématiquement dans une opposition.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Je vous propose de donner la parole à Mme Faure-Muntian, afin qu'elle nous fasse part des questions posées par les internautes au cours de cette première audition.

**Mme Valéria Faure-Muntian, députée.-** Les questions des internautes s'articulent globalement autour de quatre thèmes. Elles concernent évidemment les données, mais aussi les métiers, les infrastructures et les outils, ainsi que les algorithmes eux-mêmes.

Concernant les données, les questions se concentrent essentiellement sur l'application du règlement général de protection des données personnelles et sur les garanties apportées aux citoyens dans le cadre du SNDS et du *Health data hub*. Les internautes se demandent en outre si le traitement de ces données pourrait s'effectuer dans le cadre d'une coopération public – privé et s'interrogent, dans ce cas, sur la protection de leur vie privée. Quelles garanties de sécurité et d'anonymisation peut-on fournir, notamment pour ce qui concerne le stockage dans les *data centers*, alors que le *Cloud act* américain est potentiellement menaçant sur le sujet ? Comment garantir la qualité des données récoltées afin de pouvoir entraîner les algorithmes ? Les internautes se demandent en outre si la forme du GIP est la forme la mieux adaptée pour la gestion de ces données. *Quid* des données cliniques françaises et européennes pour l'entraînement des algorithmes ? N'est-il pas dangereux de rendre le pseudonymat irréversible ? Que faire alors si un risque sanitaire grave apparaissait lors d'une étude sur ces données ?

Les internautes souhaitent en outre connaître les évolutions induites par ces changements dans les métiers des professionnels de santé. Les algorithmes d'intelligence artificielle vont-ils répondre aux problématiques de déserts médicaux ? Permettront-ils une amélioration immédiate offrant à nos concitoyens la possibilité d'être mieux soignés ? Comment recruter les professionnels correspondant aux besoins de demain ? Les internautes se demandent par ailleurs si les médecins de la santé publique ne seraient pas les mieux à même de gérer et de gouverner le *Health data hub* ainsi que la base de données.

Sur la partie infrastructures et outils, on sait parfaitement que l'usage de l'intelligence artificielle nécessite une importante capacité de stockage et une grande puissance de calcul. Une question concerne les moyens mis en œuvre pour disposer de la puissance suffisante. Peut-on y associer les détenteurs de matériel, publics ou privés, en vue d'augmenter la capacité de puissance ? Peut-on associer la société civile au traitement des données et avec quels types d'outils ?

Concernant les algorithmes eux-mêmes, la question s'adresse essentiellement à la CNIL : l'algorithme entraîné sur des données personnelles, même pseudonymisées, entre-t-il dans le cadre réglementaire relatif aux données personnelles et à leurs usages ? Quel sera le cadre légal, juridique et éthique appliqué à cet algorithme lorsqu'il aura passé le cap de l'apprentissage et sera en fonctionnement pour traiter les patients de manière individuelle ?

**M. Cédric Villani, député, premier vice-président de l'Office.-** Voici d'excellentes questions, allant au fond du sujet.

**Mme Isabelle Zablitz-Schmitz.-** Le cadre dans lequel on se place aujourd'hui montre bien qu'il existe à la fois une dimension organisationnelle, réglementaire, technologique et financière. Dans ce contexte, il nous semble important que la gouvernance mise en œuvre laisse sa juste place à chacun des acteurs dont nous avons vu combien ils étaient variés, à l'image des sources de

données. Nous sommes dans un système où chacun a une pierre à apporter à l'édifice. Si l'État doit garantir le cadre, les règles du jeu et peut-être en partie certaines finalités pour éviter des déviations, il nous semble que la co-construction est vraiment la façon la plus intelligente de procéder, afin que la France reste dans la course et s'y positionne même avec ambition. Ceci ne pourra se faire si une partie des acteurs est exclue et si l'on ne laisse pas chacun d'entre eux jouer le rôle qui lui incombe. Par exemple, il ne faut pas que l'État se substitue à certains acteurs privés, car ceci se ferait au détriment de la science et des citoyens, dans la mesure où cela engendrerait non seulement une perte d'énergie, mais aussi de temps et surtout une perte économique. Il me paraît important que la gouvernance soit en phase avec ces éléments afin d'être efficace.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Ce thème de la nécessaire coopération souple et efficace entre partenaires publics et privés est revenue comme le fil rouge de ces auditions.

**M. Thomas Dautieu.-** Je n'aurai pas la prétention de répondre à toutes les questions liées à la protection des données, mais puis néanmoins apporter quelques précisions à ce propos. Évidemment, toutes les données non anonymes entrant dans l'écosystème dont nous avons parlé sont protégées par le règlement général sur la protection des données. En termes de garantie concernant le SNDS, la loi prévoit que les personnes soient informées par celui qui va collecter la donnée et disposent à partir de là de droits, dont celui de s'opposer à ce que cette donnée soit versée dans le SNDS. Dans le cadre de son avis sur le projet de loi, la CNIL a proposé que la future plateforme des données de santé, qui a une mission d'amélioration de la circulation de la donnée, ait aussi une mission complémentaire de meilleure information des personnes et peut-être de centralisation des droits, ce qui serait le pendant de la logique de meilleure utilisation des données.

Sur la sécurité des *data centers*, il existe un mécanisme français qui contraint, lorsque l'on héberge des données de santé, à faire appel à des hébergeurs de données de santé certifiés.

Sur les algorithmes enfin, il faut savoir qu'ils sont, en tant qu'outils de traitement des données, soumis au RGPD. En termes juridiques, les personnes ont des droits, notamment par rapport à des opérations de profilage. La CNIL a aussi réfléchi à la partie éthique des algorithmes et aux conséquences de leur utilisation, notamment en matière de santé. Elle a ainsi rendu un rapport mettant en avant les principes de loyauté et de vigilance.

**M. Erik Boucher de Crèvecœur, ingénieur expert à la direction des technologies et de l'innovation, CNIL.-** La question a été posée de savoir si les algorithmes entraînés étaient eux-mêmes des données à caractère personnel lorsqu'ils étaient ensuite mis en oeuvre. Cette question nous est souvent soumise par des *start-up* qui travaillent dans le secteur. Pour faire simple, il faut savoir que l'algorithme apprenant va catégoriser, détecter de grandes tendances, mais aussi des

*outsiders*, des cas atypiques qui vont aussi générer des catégories. A travers ces catégories, reposant sur une ou deux personnes, se pose le risque de réidentification de ces personnes. Un travail préalable doit donc être mené, consistant à nettoyer les arbres de décision, afin d'en retirer les cas trop atypiques susceptibles de donner lieu à une réidentification.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Voici quelques années, s'était produit dans ce domaine un cas assez spectaculaire : Latanya Sweeney et son équipe étaient en effet parvenus à réidentifier le dossier du gouverneur du Massachusetts. Etait-ce lié à une pseudonymisation insuffisamment forte des données ou à une faille dans la gestion par rapport aux cyberattaques ? Les nouvelles procédures permettront-elles d'éviter des situations de ce type ?

**M. Erik Boucher de Crèvecoeur.-** Je n'ai pas en tête le détail de cette affaire. Souvent, les chercheurs qui travaillent sur la réidentification de données de santé utilisent des croisements de données, des recoupements avec d'autres bases administratives ou des dossiers patients. Les réseaux sociaux fournissent également énormément d'éléments permettant de réidentifier des personnes.

**M. Jean-Marc Aubert.-** En termes de compétences, des formations doivent être développées, qui sont prévues en partie dans le plan général sur l'intelligence artificielle. Il faut en effet accroître la formation de personnes capables de traiter les données. Il ne suffit pas de disposer de l'infrastructure physique et de données très larges et diverses, dont certaines issues de l'expérience patients ; encore faut-il que les personnels de santé disposent des compétences suffisantes pour maîtriser les aspects d'utilisation de la donnée. Il ne s'agit pas de faire de tout médecin un *datascientist* ; ceci n'aurait aucun sens. Il faut seulement faire en sorte que les professionnels de santé, mais aussi les gestionnaires des établissements de santé, aient une compréhension des enjeux liés à ces données, de leurs limites et de leur intérêt. Il y a là un effort de formation important à accomplir.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Nous allons à présent devoir clore cette première table ronde. Je vous remercie pour la qualité de vos interventions.

## **DEUXIÈME TABLE RONDE : DES DONNEES AUX SOINS ET DES SOINS AUX DONNEES**

**Présidence : M. Cédric Villani, député, premier vice-président de l'Office.**

**M. Cédric Villani, député, premier vice-président de l'Office.-** Cette seconde table ronde va nous permettre de nous plonger plus encore au cœur du

sujet, au travers des aller-retour entre usages et finalités médicales, diagnostiques et thérapeutiques, de la collecte des données de santé, entre données et soins.

Nous accueillons pour ce faire un très beau *panel* d'experts, que j'ai eu l'occasion, pour la plupart d'entre eux, de rencontrer au cours des dernières années.

Nous allons commencer par écouter quelqu'un qui est devenu un proche et avec lequel j'ai édité récemment un ouvrage sur le thème. Il s'agit de Bernard Nordlinger, chirurgien oncologue à l'hôpital Ambroise Paré, membre de l'Académie nationale de médecine. Cher Bernard, un grand merci pour votre présence ce matin. Vous avez la tâche importante et difficile de donner votre point de vue de médecin sur l'utilisation des données en intelligence artificielle pour le diagnostic et la thérapeutique. Peut-être des notions telles que celle des grandes cohortes épidémiologiques en population générale feront-elles partie de votre propos. En tout cas, nous comptons sur vous pour exprimer tout ce que vous estimerez utile de porter à notre connaissance pour faire entendre le point de vue du grand professionnel de santé que vous êtes.

**M. Bernard Nordlinger, chirurgien oncologue, Hôpital Ambroise Paré, membre de l'Académie nationale de médecine.-** Merci de m'avoir invité.

L'approche que je vais vous présenter est celle d'un clinicien. Je suis chirurgien cancérologue et arrive donc au bout de la chaîne pour appliquer, de manière parfois irréversible, les options retenues préalablement du point de vue diagnostic et thérapeutique. J'ai aussi une expérience en matière de recherche clinique.

Mon exposé sera divisé en deux parties, correspondant à l'intitulé de cette table ronde. Il ira donc des données aux soins, puis des soins aux données.

Pour ce qui est de la partie allant des données aux soins, il faut entendre la notion de données au sens large du terme, c'est-à-dire incluant également les algorithmes. Il s'agira en fait d'analyser l'influence ou les applications de l'intelligence artificielle en matière de soins. Vous me pardonnerez l'aspect catalogue de ma présentation : il n'est en effet pas possible, en quelques minutes, d'entrer dans le détail. Je vais néanmoins essayer d'évoquer les principales applications, sachant que la santé est de toute évidence l'un des principaux domaines d'application de l'intelligence artificielle et de l'utilisation des données massives. Ce mouvement étant par ailleurs absolument irréversible, il va falloir faire avec.

Ceci pose avant tout un défi autour de la confiance. Il ne suffit pas en effet de disposer de réalisations techniques : encore faut-il que nos concitoyens aient confiance en ces dispositifs. Or un certain nombre de fantasmes s'expriment dans les médias : il se dit par exemple que l'intelligence artificielle ferait courir le

risque que les ordinateurs ou les algorithmes prennent la main et réduisent le cerveau humain en esclavage. Si l'on veut que le recours à l'intelligence artificielle et aux données soit bien accepté par nos concitoyens, il faut expliquer clairement qu'il s'agit d'une façon rapide et logique d'effectuer des calculs et que l'objet n'est absolument pas de réduire en esclavage le cerveau humain.

Les données sont partout, dans des bases publiques et privées. La plupart d'entre elles sont d'ailleurs stockées dans des bases privées. Une estimation récente indique en outre qu'environ 30 % des données mondiales sont en Chine, 30 % aux États-Unis et les 40 % restants partagés entre les autres pays du monde. C'est une réalité. Il faut savoir par ailleurs que tous les géants de l'internet s'intéressent à la santé, qui est l'un des principaux domaines d'application.

Quelles sont les applications, pour un clinicien comme moi, de l'utilisation de ces données pour les soins ?

La première, qui intéresse le plus le public, consiste en une aide à l'établissement du diagnostic et à la définition du traitement. Il convient de considérer que ceci ne sera jamais qu'une aide pour le médecin et ne le remplacera pas. Ces outils permettront de bénéficier d'un médecin augmenté par les informations fournies, non d'avoir un médecin automatique. En effet, un ordinateur, un algorithme, n'ont pas les compétences d'un médecin, pas plus qu'ils n'ont le sens de l'empathie ou de la vie réelle d'un patient, c'est-à-dire de sa famille, de ses antécédents, de ses bonheurs et de ses malheurs. Ils ne disposent pas non plus, qualité essentielle pour un médecin, du bon sens. Pour l'instant, les aides au choix du diagnostic ne fonctionnent pas très bien. La plus connue en oncologie est Watson, développée par IBM. Sans doute des progrès vont-ils encore être accomplis dans ce domaine. Des entreprises françaises ont par ailleurs développé des aides au choix des rendez-vous, qui fonctionnent bien.

Dans un avenir proche, des aides au choix du diagnostic seront accessibles au public afin, comme d'aucuns l'affirment, que chacun puisse être informé et prendre en charge sa santé. Il va falloir que les professionnels de santé s'habituent à faire avec. S'il est normal d'informer, il sera toutefois difficile de se passer d'un médecin. Peut-être savez-vous qu'en Chine, un petit robot passe des examens aussi bien que des étudiants en médecine. Il ne viendrait toutefois à personne l'idée de confier sa santé à un étudiant en médecine fraîchement émoulu ; mieux vaut attendre qu'il ait acquis un peu d'expérience. Un géant de l'internet, baptisé Baidu, propose par ailleurs, pour pallier le manque de généralistes, une application dans laquelle on entre ses informations et qui, à partir de leur analyse, propose un diagnostic. Bien entendu, ceci soulève de nombreux problèmes. En cas de difficulté, la responsabilité de l'acte doit rester au médecin, sauf bien entendu dans le cas d'un défaut avéré du dispositif.

Lorsque le médecin s'aidera d'un algorithme d'apprentissage, il sera en outre important que le mode de fonctionnement, de raisonnement soit explicité au

patient, afin d'éviter le phénomène de « boîte noire ». Il est essentiel de pouvoir expliquer.

Parmi les autres applications, citons la télémédecine et le télédiagnostic, qui doivent être développés afin d'améliorer l'efficacité et l'efficience des consultations et proposer une solution au problème des déserts médicaux. Bien entendu, l'organisation territoriale des soins devra par ailleurs progressivement tenir compte de la numérisation.

Il est bien évident que les futurs médecins doivent être formés aux nouvelles technologies dans les facultés de médecine. Il s'agit là d'un vaste sujet.

De nombreuses spécialités médicales vont être impactées par le développement de l'intelligence artificielle, notamment celles faisant appel à des images. On pense par exemple à la radiologie : l'IA ne remplacera pas les radiologues, mais ces derniers devront apprendre à l'utiliser. Ceci impactera aussi les métiers d'anatomopathologiste, de dermatologue, de chirurgien ou d'ophtalmologue. David Gruson a évoqué précédemment une application approuvée par la FDA américaine pour le diagnostic de rétinopathie diabétique. Il existe par ailleurs, dans le domaine de la chirurgie, des télémanipulateurs, mais aussi des dispositifs de simulation. Les oncologues et les psychiatres verront également leur pratique professionnelle modifiée par le développement de l'intelligence artificielle.

Les métiers non médicaux de l'hôpital seront aussi concernés par les nouvelles technologies. Certains vont disparaître, d'autres être créés, d'autres encore devoir s'adapter. Il me semble en tout cas que tous les métiers ayant un contact direct avec le patient ne devront pas être supprimés, mais évoluer pour appliquer les nouvelles méthodes.

Les plateformes et les tumorothèques, c'est-à-dire les bases de données permettant de corréler, de croiser les données cliniques d'un patient et de les apparier à sa maladie, doivent être développées. Il s'agit certainement de l'avenir de la recherche médicale, en tout cas en oncologie, puisque les méthodes actuelles d'essais cliniques de phase 1, 2 et 3 s'essoufflent, en raison de leur coût considérable et du grand nombre de patients nécessaires pour les mener.

Sur le thème « des soins aux données », je n'ai pas grand chose à ajouter à ce qui a déjà été dit. J'insisterai simplement sur la quasi impossibilité d'avoir une réglementation dure, rapidement dépassée par le progrès et qui risque de créer des inégalités d'accès aux soins, certains patients se rendant à l'étranger pour bénéficier de technologies indisponibles en France, ainsi que l'a expliqué David Gruson.

Il est par ailleurs important d'uniformiser les langages. Au sein de l'Assistance publique par exemple, certains hôpitaux ne parlent pas le même langage informatique.



Il faudrait également que le dossier médical partagé informatisé et accessible aux patients entre en application.

Il m'a, pour terminer, été demandé d'évoquer brièvement la cohorte Constance. Il s'agit d'une application pratique du SNIIRAM, qui a pour l'instant permis d'inclure 170 000 sujets volontaires, avec leur accord. Ces personnes sont suivies prospectivement par des interrogatoires et des examens, complétés par les données du SNIIRAM. L'intérêt en est évident pour mener à l'avenir des études épidémiologiques et de santé publique.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, M. Nordlinger.

Je vous propose d'enchaîner directement avec l'intervention suivante et d'entendre M. Emmanuel Bacry, directeur de recherche au CNRS, université Paris Dauphine, professeur et responsable de l'initiative *Big Data* de l'École polytechnique et d'un partenariat de recherche entre cette école, bientôt Institut polytechnique de Paris, et la CNAM autour de l'analyse des données du SNDS. M. Bacry est l'un des acteurs les plus respectés en France dans le domaine de l'intelligence artificielle en lien avec la santé. J'ajoute qu'il fut mon enseignant à l'École normale supérieure il y a un certain temps.

**M. Emmanuel Bacry, directeur de recherche au CNRS, université Paris Dauphine, professeur et responsable de l'initiative *Data Science*, École polytechnique.-** Je vais vous parler brièvement de ce partenariat, qui a commencé voici quatre ans entre la CNAM et l'École polytechnique et dont le but est de tester les technologies *big data* ou d'intelligence artificielle, c'est-à-dire les algorithmes, sur les données du SNIIRAM.

Lorsque cette initiative a été lancée, le SNDS n'existait pas, mais il faut savoir que les données du SNIIRAM correspondent à peu de choses près à celles du SNDS. Notre partenariat ne s'inscrit donc pas dans le cadre du SNDS. Nous intervenons dans une bulle sécurisée, totalement coupée du monde extérieur, avec accès à cette pépite assez unique qu'est le SNIIRAM, l'une des plus grosses bases de données médico-administratives au monde, que nombre de pays nous envient.

Le SNDS représente à mon avis un grand progrès, puisqu'il a énormément facilité l'accès aux données. On peut ainsi désormais obtenir en quatre mois une réponse sur l'acceptation ou le refus d'un accès aux données pour un projet particulier. De nombreux projets passionnants, avec un très fort impact, existent grâce au SNDS.

Il convient toutefois de préciser très clairement que tout ce qui relève des algorithmes modernes d'intelligence artificielle ne peut pas tourner sur l'infrastructure telle qu'elle existe aujourd'hui dans le SNDS. Si de nombreuses utilisations sont rendues possibles par cette infrastructure, d'autres en revanche,

pourtant extrêmement importantes, ne sont pas accessibles. Il s'agit là de la première difficulté à laquelle nous nous sommes trouvé confrontés dans ce partenariat, qui nous a conduits à envisager de mettre en place les premiers éléments d'une restructuration d'un pipeline *big data* sur les données du SNIIRAM ou du SNDS. Ceci représente un très gros travail. Nous avons ainsi mis deux ans, avec en vitesse de croisière près de dix personnes à temps plein (des chercheurs, des experts en données et cinq développeurs) pour poser les premières briques de cette infrastructure et commencer à pouvoir travailler sur des algorithmes en IA. Ce travail a consisté en une refonte de l'infrastructure machine, de la base de données elle-même et de l'accès à cette base, avec des bibliothèques permettant d'accéder de façon efficace à la base, pour construire des données susceptibles d'être fournies à des algorithmes d'intelligence artificielle, afin de disposer d'outils accessibles d'intelligence artificielle moderne, permettant de faire tourner ces algorithmes sur cette base.

Nous continuons à œuvrer à l'amélioration de ces premiers éléments, mais avons également réellement commencé à faire de l'intelligence artificielle. Nous avons par exemple travaillé sur des problèmes de pharmacovigilance. Nous nous sommes notamment demandé si nous serions capables de concevoir un algorithme en mesure d'identifier automatiquement, ou en tout cas avec le moins d'intervention humaine possible, des médicaments qui sont sur le marché et présentent potentiellement des effets néfastes secondaires. Nous avons travaillé sur ce type d'algorithme et rencontré un premier succès fin 2017, avec la mise au point d'une première version. Il était évidemment hors de question de le tester sur un cas réel. Pour autant, il nous fallait apporter la preuve que le système fonctionnait. Nous avons donc ré-identifié, avec des méthodes totalement différentes de celles employées habituellement, un médicament parmi un groupe de médicaments retirés du marché en 2011 pour cause d'effets secondaires, en l'espèce de risque de cancer de vessie aggravé.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Qu'entendez-vous dans ce cas par « réidentifier » ?

**M. Emmanuel Bacry.-** Le médicament avait été identifié par les équipes scientifiques de la CNAM, par des algorithmes de survie, qui sont des algorithmes standards pour effectuer de l'identification, avec validation d'hypothèses. On pose une question précise (est-ce que tel médicament pose tel problème ?), puis des experts se réunissent pour savoir ce que signifie être exposé à ce médicament. Ceci nécessite notamment un travail de préparation des données et peut prendre plusieurs mois. L'idée ici est de développer un algorithme qui soit automatisable, nécessite peu d'interventions d'experts et mette des coups de projecteur sur des médicaments qui devront ensuite suivre le *pipeline* standard de biostatistique, qui est l'état de l'art aujourd'hui.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Dans le premier cas, si nous comprenons bien, on a une suspicion laissant penser qu'un

médicament donné provoque tel effet et l'on pose la question correspondante. Dans le second cas, on « part à la pêche ».

**M. Emmanuel Bacry.-** C'est cela. Aujourd'hui, nous « partons à la pêche » sur une situation réelle, avec plus de 400 médicaments. L'effet secondaire que nous essayons d'identifier est le suivant : existe-t-il, parmi ces médicaments, des médicaments provoquant une augmentation du risque de chute pour les personnes âgées ? Nous faisons tourner l'algorithme et le modifions. Je rappelle qu'il s'agit de travaux de recherche.

Nous travaillons aussi sur les parcours de soin, en vue de leur optimisation. Nous essayons ainsi de faire de l'IA pour structurer les différents parcours de soin au sein d'une pathologie. Ceci permettrait par exemple de connaître les parcours-types, de distinguer ceux qui soignent le mieux, ceux qui coûtent le moins cher. Il y a là des enjeux économiques extrêmement importants.

Je voudrais profiter des quelques minutes qui me restent pour faire le lien avec certaines questions soulevées lors de la première table ronde. Il était très clair pour nous dès le début de ce projet que nous construisions un *pipeline* qui devait servir de socle pour une ouverture plus large, encadrée juridiquement et sécurisée, du SNDS. Nous souhaitions partager cette ressource. J'ai ainsi été, dès l'origine du projet de *Health data hub*, l'un de ses plus ardents défenseurs. Il faut bien se rendre compte que si l'on parvient à mettre dans une infrastructure moderne ne serait-ce que les données du SNDS et les données cliniques des hôpitaux publics – le *hub* a en effet vocation à faire plus que cela – alors on disposera sans aucun doute de l'une des plus belles bases de données de santé au monde. Il y aura là un rayonnement mondial assuré : tout le monde (industriels, académiques) va vouloir venir travailler en France. Il s'agit d'un élément important à comprendre.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Concernant les données cliniques, s'agit-il de données déjà existantes, formatées, ou est-il question d'aller jusqu'aux comptes rendus d'opérations, aux analyses de textes ?

**M. Emmanuel Bacry.-** Il faut évidemment aller le plus loin possible. On peut déjà, avec les données cliniques organisées dans certains hôpitaux, aller extrêmement loin, avec un impact fort. Il n'est pas nécessaire d'attendre une réorganisation totale de tous les entrepôts de données de tous les hôpitaux pour agir.

Je terminerai en évoquant un point qui m'apparaît particulièrement important. Il a été question, à propos de ce *hub*, de l'exigence d'avoir une structure agile. Je connais bien le monde académique, ainsi que celui des *start-up* et de l'industrie en IA. Il faut que cette structure soit attractive et l'on ne parle pas que d'argent. Il faut bien évidemment pouvoir payer correctement les *datascientists* et les ingénieurs ; mais cette attractivité résidera aussi dans le fait de pouvoir dire aux

gens qu'ils trouveront là une animation scientifique et que les plus grands chercheurs en IA, français et internationaux, collaborent avec le *hub*. C'est la perspective de travailler avec des chercheurs de haut niveau sur des sujets présentant des enjeux sociétaux extrêmement importants, que ce soit avec l'univers public ou la sphère industrielle, qui va attirer les meilleurs talents de toutes les nationalités.

On sait aujourd'hui que les GAFAs sont arrivés en France, notamment dans le domaine de la santé. Ceci effraie beaucoup de gens ; je pense pour ma part qu'il s'agit d'une très bonne nouvelle, car cela signifie qu'il se passe en France actuellement des choses très intéressantes en matière de santé. En revanche, il est évident qu'il faut organiser un contre-pouvoir à ces GAFAs. L'État n'y parviendra pas en termes financiers. Par contre, ceci peut s'effectuer dans le cadre de l'organisation des données, en mettant des conditions d'accès et en ayant une base de données bien organisée, à laquelle tout le monde pourra avoir accès. Je pense que ceci constituerait un contre-pouvoir extrêmement important.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, M. Bacry. L'organisation de ces contre-pouvoirs suppose une élaboration dans la discussion et non dans l'antagonisme. Ceci explique que nous entendrons tout à l'heure la voix d'un représentant de l'un de ces fameux GAFAs, que nous connaissons bien par ailleurs dans le domaine de la recherche française.

Nous allons pour l'instant donner la parole à l'Inria, l'un des acteurs majeurs de l'intelligence artificielle en France, opérateur en particulier de la réorganisation du système de recherche en la matière, et plus précisément à l'un de ses directeurs de recherche, M. Nicholas Ayache, de l'équipe-projet Epione (e-patient et e-médecine), qui va évoquer en détails certains aspects plus ou moins techniques de constitution, stockage, interrogation et identification de données en lien avec la santé.

**M. Nicholas Ayache, directeur de recherche à l'Inria, équipe-projet Epione (e-patient et e-médecine).-** Mon exposé va aborder plus spécifiquement la question de l'imagerie médicale à l'heure de l'intelligence artificielle.

On l'a vu, l'intelligence artificielle permet d'augmenter les capacités cognitives des médecins, pour les aider à mieux prendre en charge leurs patients. Appliquée à l'imagerie médicale, elle permet d'analyser des images très complexes, en prenant en compte des masses de données considérables qu'aucun cerveau humain ne peut appréhender dans leur totalité.

Des exemples récents ont frappé les esprits. J'en citerai trois. En dermatologie, un logiciel d'apprentissage profond, d'abord entraîné sur plus d'un million d'images naturelles, a ensuite été ajusté sur les images de plus de 100 000 lésions dermatologiques, pour apprendre à distinguer automatiquement les lésions cancéreuses des lésions bénignes, comme un dermatologue expert.

Le deuxième exemple concerne le secteur de la radiologie : l'entreprise française Therapixel, représentée ici par son fondateur Olivier Clatz, qui interviendra tout à l'heure, a entraîné un logiciel d'apprentissage sur plus d'un demi-million de mammographies dans le cadre d'un défi mondial et remporté la compétition, en distinguant les mammographies suspectes des mammographies normales mieux que tous ses concurrents. Le logiciel se compare aujourd'hui aux meilleurs radiologues experts sur ce sujet.

En ophtalmologie enfin, des logiciels d'apprentissage entraînés sur plus de 100 000 images de la rétine sont désormais capables de détecter les rétinopathies diabétiques aussi bien qu'un ophtalmologue expert, à tel point que, comme l'a indiqué M. Gruson précédemment, la FDA, agence américaine du médicament, a pour la première fois dans son histoire autorisé un tel logiciel à faire ce diagnostic automatiquement, sans que l'image soit regardée par un ophtalmologue.

Il faut savoir que tous ces travaux autour des images médicales ne se limitent pas au développement de logiciels d'apprentissage profond, qui posent par ailleurs d'autres problèmes dans la mesure où ils s'appuient sur des millions de paramètres difficiles à interpréter ; leur ajustement requiert ainsi d'immenses bases de données renseignées, qui coûtent très cher et sont souvent difficiles à acquérir. Il est aujourd'hui possible de construire des modèles numériques du patient, qui exploitent les connaissances dont nous disposons en anatomie et en physiologie, pour se limiter à un nombre réduit de paramètres plus facilement interprétables. Il s'agit en fait d'algorithmes avec des modèles mathématiques, physiques, biologiques, qui permettent de construire, à partir des images médicales et d'autres données sur le patient, une représentation numérique et personnalisée de ce patient, que l'on qualifie de « patient numérique personnalisé », de « e-patient » ou encore de « jumeau numérique ». Les paramètres et les modèles de ce « jumeau numérique » peuvent ensuite être utilisés par des algorithmes de médecine numérique, c'est-à-dire d'aide au diagnostic, au pronostic et à la prise en charge thérapeutique.

Ces modèles biophysiques du patient numérique permettent aussi d'augmenter les bases de données, en simulant par exemple des maladies rares ou des cas peu ou mal représentés dans les bases de données déjà acquises. Ils permettent donc de faire de la complétion de bases de données.

Enfin, le cadre théorique de ce patient numérique et les méthodes modernes des sciences des données permettent d'intégrer des données sur le patient que l'on pourrait qualifier de « holistiques », c'est-à-dire incluant non seulement les images anatomiques et fonctionnelles du patient, mais aussi les données biologiques (génétiques, métaboliques, etc), voire des données comportementales ou environnementales, sur le style de vie par exemple. On est ensuite en présence de données extrêmement hétérogènes, de grande dimension : il s'agit véritablement d'un défi scientifique que de les analyser. La recherche est très active sur ces sujets, notamment dans le cadre des Instituts interdisciplinaires

d'intelligence artificielle en cours de déploiement en France. La date limite de dépôt des projets était il y a deux jours. Ceci va permettre de faire progresser la compréhension de certaines maladies, de mieux guider leur prise en charge. On pense notamment au dépistage et au traitement de nombreux cancers, dont ceux du poumon, du sein ou de la prostate.

À terme, le champ d'application de l'intelligence artificielle en imagerie médicale est appelé à servir un très grand nombre de disciplines médicales, dont la radiologie, l'oncologie, mais aussi la cardiologie, la neurologie, la radiothérapie, la chirurgie, etc.

Que se passe-t-il à l'étranger ? Au Royaume-Uni par exemple, la base de données UK Biobank met à la disposition des chercheurs l'ensemble des données (données d'images, génétiques et environnementales) acquises sur plus de 100 000 participants volontaires, suivis dans le temps. Il s'agit donc d'une base de données dynamique.

Aux États-Unis, la base de données ADNI propose, pour une immense cohorte de patients atteints de la maladie d'Alzheimer, des images du cerveau au cours du temps et des données cliniques et génétiques.

En France, des initiatives récentes ont été lancées pour constituer des entrepôts de données de santé associant de l'imagerie et des informations complémentaires sur les patients, afin de mettre ces données à disposition des chercheurs, en préservant leur confidentialité et avec l'accord des participants. A été évoqué précédemment le *Health data hub*, mais on pourrait citer également l'entrepôt de données de santé de l'AP-HP, France Life Imaging ou encore DRIM France IA, dont Laure Fournier va nous parler dans quelques instants.

Je pense qu'il est nécessaire, concernant ces bases de données, de rappeler l'importance de s'assurer de la bonne représentativité des cas rares, qui par définition ne sont pas nombreux, de l'absence de biais liés au genre, à l'âge, à l'ethnicité et de la présence de sujets sains en nombre suffisant lorsque l'on souhaite effectuer des comparaisons (ce qui pose un problème éthique lorsque la modalité d'imagerie est irradiante par exemple).

Il faut selon moi vraiment encourager ces initiatives, indispensables pour permettre à la recherche de faire des progrès considérables, et veiller à ce que la réglementation, si elle protège suffisamment les patients, ne pénalise pas les chercheurs français et européens face au reste du monde.

Je souhaiterais insister pour terminer sur le fait que l'intelligence artificielle, l'imagerie médicale et le patient numérique sont un ensemble d'outils informatiques au service de la médecine dite « des 4 P » (personnalisée, précise, prédictive, préventive), destinée à mieux soigner non pas le patient numérique, mais bien le patient réel.

Pour faire écho aux propos de Bernard Nordlinger, je voudrais conclure sur le fait que ces nouveaux outils ont vocation à aider le médecin, non à le remplacer. Un certain nombre de qualités du médecin, comme le bon sens, l'empathie, la compassion, la compréhension globale du patient, l'esprit critique, la conscience professionnelle et la créativité, restent encore et pour longtemps l'apanage de l'intelligence humaine et demeurent absolument indispensables dans ce domaine.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, M. Ayache. Il est impressionnant de constater que, sur certains de ces sujets hautement techniques, la reconnaissance par imagerie automatique arrive à faire sensiblement aussi bien que les meilleurs experts humains. Le fait que la FDA ait approuvé l'usage de l'un des ces outils de diagnostic sans regard humain est assez emblématique. On pourra se demander s'il existe un précédent dans la méthodologie de la FDA en la matière, ayant vocation à essaimer.

Nous accueillons à présent le professeur Laure Fournier, spécialiste de radiologie, membre de la Société française de radiologie, qui enseigne cette discipline. La radiologie est régulièrement évoquée comme une profession appelée à évoluer de façon considérable au regard des progrès rapides effectués en imagerie. Certains évoquent la disparition des radiologues, d'autres estiment que cette évolution ouvrira une nouvelle ère pour cette spécialité, faisant appel à de nouvelles compétences. Que pensez-vous de tout cela, Mme Fournier ?

**Mme Laure Fournier, PU-PH, Société française de radiologie, Collège des enseignants en radiologie de France.-** Merci pour cette introduction. À vrai dire, je préfère me voir comme un phénix qui va renaître des cendres auxquelles certains réduisent, peut-être un peu hâtivement, l'avenir de la radiologie.

Je suis très honorée de représenter ici la communauté des médecins radiologues et vais essayer, dans le temps court qui m'est imparti, d'aborder quatre points qui nous paraissent essentiels.

Le premier concerne l'importance des bases de données. Il s'agit, comme ceci a été souligné à plusieurs reprises, d'un enjeu national, scientifique et industriel, mais aussi médical pour le patient. En effet, les algorithmes qui sont développés et vont conduire à des innovations thérapeutiques ou à des évaluations de nos pratiques de soin sont intrinsèquement dépendants de la population sur laquelle ils ont été développés. Si l'on se contente par conséquent d'utiliser des algorithmes entraînés sur des populations américaines, chinoises ou indiennes, ils ne s'appliqueront pas forcément aux patients français (qui ont des profils de maladies différents), à nos habitudes de soin et à nos politiques de santé. Il conviendra donc, au minimum, de les valider sur nos propres populations.

Au-delà de ces bases de données au sens général, j'insiste également sur l'importance des bases de données en imagerie et le fait de disposer des images

elles-mêmes. Lorsque j'ai parlé de mon métier à un ami, il m'a dit que l'on pouvait finalement le résumer par la formule « voir c'est savoir et savoir c'est pouvoir ». J'ai trouvé cette phrase très belle. Il est vrai que les images sont au cœur du soin. Je travaille en cancérologie et bien souvent les images permettent de voir la maladie pour la première fois, de savoir si elle est localisée afin de la confier aux chirurgiens comme M. Nordlinger ou au contraire si elle est malheureusement plus étendue et doit être confiée aux cancérologues. Je pense par ailleurs que l'opportunité de ces bases de données en imagerie va au-delà de la maladie qui nous intéresse au moment du soin, car les images apportent une information bonus, en permettant de voir tous les autres organes. Ceci est notamment intéressant pour acquérir des connaissances sur l'histoire des maladies, sur les maladies chroniques, comme la bronchopathie chronique, l'état de nutrition du patient, l'ostéoporose, etc. Il s'agit d'une source d'information extrêmement riche, à condition de pouvoir disposer des images elles-mêmes. Ces données sont parfois insoupçonnées au moment du soin, ne sont pas explicitement détaillées, mais sont une mine d'information et l'une des grandes opportunités de ces bases.

Le Conseil professionnel de la radiologie, qui réunit les radiologues libéraux, hospitaliers et universitaires, conduit une initiative intitulée DRIM France IA, qui vise à regrouper toutes les images de soin dans une base de données, pour servir en partie d'interface au *Health data hub* et pouvoir ainsi rapprocher ces images, qui n'ont pas de sens seules, des données cliniques.

Comment passer du soin aux données ? Il s'agit de la première étape. Je souhaiterais rappeler ici qu'il convient de ne pas oublier l'investissement initial sur ces données. Je suis d'accord avec Emmanuel Bacry sur l'idée que l'on peut faire beaucoup de choses avec les données telles qu'elles sont aujourd'hui, mais pense que l'on a beaucoup à gagner à investir pour s'assurer de la qualité de ces données, de leur véracité. En pratique, certaines informations sont parfois implicites, voire orales. Elles sont parfaitement compréhensibles dans le contexte, mais seront peut-être, une fois sorties du contexte et intégrées dans une base, plus difficiles à appréhender et à interpréter. Il me semble important de réfléchir à cet aspect. Il existe en outre un bénéfice secondaire, dans la mesure où dès que l'on travaille sur les données de soin et leur structuration, on améliore la qualité des soins de manière immédiate. Ceci engendre donc un bénéfice immédiat pour le patient, au jour le jour.

Comment, enfin, passer des données aux soins ? Encore une fois, il est important de se rappeler à quoi servent ces bases de données : que les acteurs soient publics ou industriels, l'objectif est d'améliorer la santé de nos patients et de notre société en général. La pertinence médicale doit donc être au cœur des travaux. Si l'on veut que ces bases soient des outils utiles, elles doivent être pertinentes. Ceci doit évidemment être assorti d'une garantie de leur validité et de leur sécurité. Je veux ici souligner la volonté des radiologues de contribuer à ce processus. Lorsque l'on a commencé à parler des bases de données et de l'intelligence artificielle, ceci



a suscité beaucoup d'appréhensions dans notre milieu. Je trouve toutefois que la situation a beaucoup évolué. Aujourd'hui, cela génère beaucoup d'enthousiasme lucide, de volonté. De nombreuses initiatives se développent, dont DRIM France IA. Je pense que nous sommes prêts, au sens où nous le souhaitons et nous y préparons, à accompagner ces développements et à servir, parmi d'autres, de garants de la qualité et de la sécurité.

Il faut tout d'abord qu'il y ait suffisamment de radiologues. Ceci reste une préoccupation, dans la mesure où nous éprouvons déjà des difficultés à couvrir les besoins du territoire pour nos missions actuelles. *Quid*, dans ce cas, de nos missions futures ? Au-delà de cet aspect, il faudra aussi disposer de suffisamment de radiologues formés. Je fais partie du Collège des enseignants en radiologie, où je suis responsable de la formation des internes en intelligence artificielle. Les cours ont déjà commencé. Nous formons ainsi les étudiants en médecine non seulement à la technique, mais avons également choisi d'associer à chaque heure de formation en technologies numériques une heure de formation en relation patient. Nous considérons en effet qu'il va désormais falloir savoir mieux communiquer aux patients les résultats issus de ces outils. Sans être mélodramatique, il faut imaginer comment recevoir un patient après son scanner pour lui indiquer que l'on a découvert une lésion sur l'un de ses reins et que l'algorithme indique qu'il y a 48 % de chances que ce soit un cancer. Il va falloir être capable de leur fournir des explications, de les accompagner, pour que ce passage de la donnée aux soins soit acceptable sur le plan humain, pour le patient comme pour le médecin.

Je vous remercie et espère que nous aurons l'occasion de rediscuter de certains de ces points.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, Mme Fournier. Nous sommes très heureux d'entendre que la formation est lancée. Il s'agit en effet d'un point clé pour l'avenir, particulièrement difficile à appréhender par la puissance publique, dans la mesure où l'enseignement est structurellement indépendant du pouvoir politique. Votre intervention va tout à fait dans le bon sens.

Continuons notre tour de table avec M. Olivier Clatz, fondateur de Therapixel, *start-up* d'édition de logiciels d'imagerie médicale, qui va nous faire part du témoignage de l'un des acteurs au cœur de la problématique de l'utilisation de l'intelligence artificielle et de la donnée pour les applications médicales.

**M. Olivier Clatz, fondateur de Therapixel, start-up d'édition de logiciels d'imagerie médicale.-** Je suis très content du chemin parcouru depuis un an et la mission dite « Villani ». Je ne pensais pas que nous aurions un jour ce type de débat à l'Assemblée nationale. Lors de mes premiers rendez-vous dans les centres hospitaliers voici quelques années pour essayer de travailler sur ces données, la situation était extraordinairement compliquée. Aujourd'hui, tout le monde est sensibilisé à ce sujet, à tous les niveaux de la pyramide administrative et

politique. C'est donc une réelle fierté pour moi que de participer à ce débat et une satisfaction de constater que les choses avancent.

Therapixel développe des logiciels pour l'assistance au dépistage du cancer du sein, sur la base de l'examen de la mammographie. Nous sommes tout d'abord passés par le parcours du combattant et notamment la phase de soumission des dossiers CNIL, CRS, SNDS. Il est très bien que tout ceci existe. Ainsi, la possibilité de dérogation permet de ne pas avoir à redemander leur consentement à toutes les patientes, ce qui parfois n'est pas possible. Par contre, lorsque l'on soumet ce type de dossier, on se trouve parfois devant le phénomène de *black box*, reproche d'ailleurs souvent fait aux algorithmes : on ignore en effet ce qu'il faut dire pour obtenir l'autorisation. On présente par conséquent le projet en expliquant précisément ce que l'on compte faire, mais il serait utile de disposer de *guidelines*, de recommandations, en particulier sur la manière d'organiser la sécurité du système afin de garantir l'absence de fuite des données.

L'organisation du dépistage du cancer du sein est très particulière en France, puisqu'il existe des structures spécifiques, qui nous avaient semblé constituer la bonne porte d'entrée pour le projet que nous souhaitions mener. Or il s'agissait d'une fausse piste, dans la mesure où ces structures fonctionnent, en 2019, avec des films, même si les données sont acquises en numérique. Il n'est donc pas possible de travailler en intelligence artificielle sur ces données. Nous avons donc décidé de contacter les centres de lutte contre le cancer, afin de trouver des patientes en cours de développement de la maladie ou avec une suspicion. La problématique est que le dépistage est aujourd'hui très peu fait dans ces centres de lutte, mais se déroule essentiellement dans des centres privés. Nous avons pensé qu'il serait possible de surmonter cette difficulté en considérant les cancers dans les centres de lutte, les sujets dans les cabinets et en rapprochant les morceaux. Or, cette démarche s'est avérée plus difficile que nous l'avions imaginée. En 2007, une décision de la CNIL a en effet interdit d'utiliser un identifiant national sur la base du NIR. Nous éprouvions donc d'énormes difficultés à associer un examen effectué dans un centre 1 à un autre examen réalisé dans un centre 2, sur la même personne. Cette situation a toutefois été modifiée par une loi de 2016, dont le décret d'application est paru en 2017. Il apparaît toutefois que cette possibilité est encore peu appliquée.

**M. Cédric Villani, député, premier vice-président de l'Office.-** De quelle possibilité parlez-vous ?

**M. Olivier Clatz.-** Je faisais référence à la possibilité d'utiliser le NIR.

**Mme Jeanne Bossi-Malafosse.-** Le numéro de sécurité sociale est devenu l'identifiant national de santé depuis la loi de 2016 et l'on a jusqu'à 2020 pour qu'il soit utilisé comme tel dans tous les systèmes d'information de santé, dès lors qu'ils participent à la prévention et à l'administration des soins. Cette reconnaissance n'était pas simple, car elle posait des problèmes importants de

protection des données, qui ont depuis lors été réglés par la CNIL. Il fallait inscrire dans la loi le fait que le NIR devienne l'INS : c'est désormais chose faite. Il reste aux organisations à le mettre en place d'ici le 1<sup>er</sup> janvier 2020.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Il semblerait donc que la situation soit réglée, sur le papier tout du moins. Il faut simplement le temps que la mise en œuvre soit effective.

**Mme Jeanne Bossi-Malafosse.-** Une personne qui souhaiterait aujourd'hui utiliser le numéro de sécurité sociale comme INS ne se fera pas retoquer par la CNIL.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Confirmez-vous ceci, M. Dautieu ?

**M. Thomas Dautieu, directeur de la conformité, CNIL.-** Absolument.

**M. Olivier Clatz.-** J'ai effectué une estimation du temps que nous passions chaque mois pour récupérer 10 000 données d'examens : ceci nous prend six hommes-mois. J'ai eu la semaine dernière un rendez-vous avec un établissement à Chicago, qui est l'un des dix plus gros systèmes de santé américains : nous avons réussi à faire en environ cinq minutes l'équivalent du travail de dix hommes-an, car les données étaient structurées, les patients avaient toujours le même numéro, l'on disposait de toutes les images, de toute l'anatomopathologie, de l'ensemble des comptes rendus. Ceci rend assurément les choses beaucoup plus faciles.

Fort de ces expériences, je me permets de vous proposer un certain nombre de recommandations.

La première consiste à s'assurer que l'on mette vraiment en œuvre l'utilisation du NIR, pourquoi pas avec un volet incitatif. Dans son rapport, Jean-Marc Aubert suggérait de modifier les règles des remboursements avec une dimension qualité. Je pense que l'un des points importants de la qualité est de s'assurer que les données sont associées aux bons patients. L'un des leviers les plus importants de l'État étant, comme l'a rappelé M. Longuet, la contrainte, il serait possible d'exercer une certaine contrainte dans ce domaine et de s'assurer que l'on dispose des bonnes données pour les bonnes personnes.

Aujourd'hui, il existe par ailleurs autant de logiciels que d'établissements de santé. Chaque cabinet, chaque hôpital a son logiciel, qui ne communique pas avec les autres. À l'heure d'internet, ceci pose problème et l'on aurait tout intérêt à centraliser au minimum la gestion des données. L'État devrait ainsi reprendre la main sur les données, sans bien entendu empêcher le développement d'applications industrielles comme celles que nous mettons en œuvre. Mon rêve est de pouvoir disposer en France d'un système tel que celui

existant aujourd'hui en Estonie. L'exemple de l'établissement de Chicago est également parlant. L'idéal serait ainsi de disposer d'un système de santé intégré, capable de rassembler toutes les données (images, données génétiques dans les limites de ce que la loi permet, données d'anatomopathologie, etc).

Certaines solutions sont en cours de mise en œuvre. Selon le rapport de la mission confiée à Dominique Pon et Annelore Coury, l'Espace numérique de santé devrait permettre cela. Je pense par conséquent qu'il importe de soutenir très fortement cette initiative, notamment financièrement. Or la loi ne précise pas où trouver les financements nécessaires en investissement et en fonctionnement, qui sont certainement de l'ordre de plusieurs centaines de millions d'euros.

À titre personnel, je m'interroge, relativement au projet de loi, sur la capacité offerte aux citoyens d'effacer leur propre dossier. Je pense en effet que, si la communauté rembourse ou soutient le système de santé et permet d'obtenir ces données numériques, alors le citoyen devrait, en retour, donner accès aux informations le concernant et ne pas pouvoir les supprimer. Il s'agit là d'un avis personnel, autour d'une question qui mériterait selon moi d'être débattue à l'Assemblée nationale.

Il faudrait bien entendu intégrer dans cet Espace numérique de santé les données massives. Il existait dans le DMP une recommandation contraire. Ce serait dommage, à l'ère du *big data*.

L'un des points majeurs est aussi de veiller à bien structurer la connaissance médicale, de la mettre dans des cases, afin de pouvoir ensuite exploiter plus facilement les données.

Comme il en a été beaucoup question précédemment, je n'insisterai pas sur l'importance de disposer d'un outil tel que le *Health data hub*, agile, capable de mener les recherches que nous avons mentionnées, dans l'intérêt public ou pour les entreprises.

L'Espace numérique de santé permettrait par ailleurs de récolter le consentement ou l'opposition, chaque patient étant informé dans ce cadre des recherches effectuées sur ses données.

Il est, pour toutes ces raisons, très important de soutenir le développement des outils que sont le *Health data hub* et l'Espace numérique de santé, en leur octroyant les moyens et toute la souplesse nécessaires.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, M. Clatz.

Mme Combes, souhaitez-vous intervenir ?

**Mme Stéphanie Combes, cheffe du projet *Health data hub*, DREES.-**

Je souscris tout à fait aux propos de M. Clatz. Nous avons commencé à discuter avec la nouvelle mission Numérique en santé, notamment avec Dominique Pon qui a été nommée à cet effet, dans le prolongement du travail sur le virage numérique. Je pense qu'il existe, dans la perspective de la mise en place de l'Espace numérique de santé, des ponts à créer entre cet Espace et le *Health data hub*, pour permettre de récolter plus facilement des données auprès des citoyens, en les informant, et d'apparier ces données *by design* plutôt que de les réapparier par la suite. Ces données, si elles sont riches et recueillies auprès de tous les citoyens, pourraient ensuite être reversées à des fins de recherche dans un outil comme le *Health data hub*. Le lien entre les deux dispositifs me semble donc essentiel.

**M. Claude Gissot, directeur de la direction de la stratégie, des études et des statistiques, CNAM.-** L'idée qu'une partie des données vienne alimenter le DMP, donc l'Espace numérique de santé, visait à faire en sorte qu'une partie nominative entre dans le DMP pour le patient et que ces mêmes données puissent partir de façon pseudonymisée dans le *Health data hub*, ce qui permettrait d'éviter la déconnexion et la décorrélation entre des données individuelles arrivant du côté DMP et des données versées dans le *Health data hub* ou le SNDS. Il s'agirait ainsi de disposer des mêmes données d'un côté et de l'autre, ce qui faciliterait les ponts en termes d'analyse des données dans le *Health data hub* et pour leur application à des sujets individuels du côté de l'Espace numérique de santé.

**M. Cédric Villani, député, premier vice-président de l'Office.-** L'intervention de M. Clatz a fourni de très précises et intéressantes recommandations, ainsi qu'un parangonnage international précieux, sur lequel il est clair que nous aurons l'occasion de nous pencher à nouveau avant l'examen du projet de loi à venir.

Nous poursuivons cette table ronde par l'exposé de M. Rémy Choquet, directeur de l'innovation d'Orange *healthcare*, chercheur associé au laboratoire d'informatique médicale LIMICS de l'INSERM, sur le sujet large du numérique à l'hôpital. Nous avons évoqué à plusieurs reprises l'importance d'associer tous les acteurs, publics et privés, dans cette aventure. Par votre position au sein d'Orange *healthcare* et votre qualité de chercheur associé à un laboratoire de l'INSERM, vous incarnez parfaitement cette association. Nous vous écoutons, M. Choquet.

**M. Rémy Choquet, directeur de l'innovation d'Orange *healthcare*, chercheur associé au laboratoire d'informatique médicale LIMICS (U1142) de l'INSERM.-** J'ai pendant longtemps travaillé dans le secteur public et y poursuit encore aujourd'hui des activités de recherche. J'ai géré des données de santé dans les maladies rares pour le compte de l'État pendant quelques années, avant de rejoindre le secteur de l'industrie privée.

Il existe plusieurs positionnements possibles. J'observe par exemple chez Orange, qui est une grande entreprise française du numérique, un positionnement,

qui peut différer de celui d'autres structures, consistant à ne jamais considérer les données de santé comme un objet de valorisation pour la société, mais comme un objet à partir duquel on va accompagner le secteur de la santé dans le développement d'une certaine industrialisation des dispositifs mis en place dans ce domaine.

Les données de santé sont des éléments susceptibles d'avoir un impact populationnel et/ou individuel, dans un processus de santé. Evidemment, ces deux visions ne se confondent pas et ne sont pas nouvelles. Des cohortes ont ainsi été mises en place. La France se situe toutefois dans ce domaine bien en-deçà de ce qui a pu être élaboré dans d'autres pays, faute d'investissements suffisants. Il n'est pas question aujourd'hui, me semble-t-il, de résoudre ce problème avec des données médico-administratives, puisque la précision et la définition des données nécessaires pour faire avancer la science sont beaucoup plus fines et de nature totalement différente de celles que contiennent les grandes bases de données médico-administratives.

La deuxième question est de savoir quelle valeur peuvent avoir les données de santé cliniques collectées lors des processus de prise en charge à l'hôpital ou par les différents professionnels de santé, dans l'optique d'une réutilisation soit pour des décisions populationnelles, soit pour des décisions cliniques précises. Cette question reste ouverte dans beaucoup de secteurs. Je pense qu'il est important d'y réfléchir globalement et de prioriser les éléments pour lesquels l'investissement public permettra de faire avancer la science d'une part, les pratiques d'autre part.

Nous répondons, chez Orange, à des appels d'offres orientés vers les hôpitaux et les collectivités territoriales, afin de mettre en place en leur sein des dispositifs numériques. Le secteur de la santé est en transformation permanente et se numérise progressivement. Il existe ainsi une corrélation directe entre la vitesse à laquelle cette numérisation intervient dans le secteur et le retour sur investissement que vont permettre les outils numériques mis à sa disposition. La dernière brique du retour sur investissement est l'analyse, la capacité qu'ont les professionnels de santé, mais aussi les directeurs d'hôpitaux, les autorités sanitaires et les pouvoirs publics au sens large à prendre les bonnes décisions grâce aux données générées.

Orange a choisi de se positionner dans ce champ en procédant au rachat de la société française Enovacom, qui distribue un entrepôt de données médicales développé au CHU de Rennes sur six ou sept hôpitaux français. L'idée est vraiment de pouvoir équiper les professionnels de santé, les hôpitaux, avec des outils leur permettant de mieux capitaliser sur ces données. Evidemment, certains enjeux dépassent le simple problème de l'hôpital ; mais l'hôpital est un maillon de notre point de vue assez important, car il est producteur de données. Même si l'on installe des cadres techniques, de l'interopérabilité, des logiciels similaires, l'interprétation d'une donnée de santé à l'extérieur de l'hôpital et du cercle des gens qui ont produit ces données peut être tout à fait différente. Il ne faut donc pas

négliger l'investissement à mettre en place dans les hôpitaux, pour que toute la chaîne de valeur autour des données de santé puisse être abordée.

Les données ont une valeur, à plusieurs niveaux et pour divers types d'acteurs. Celle-ci est difficilement modélisable. Néanmoins, *in fine*, qu'il s'agisse de surveillance épidémiologique, de diagnostic, de thérapeutique ou d'efficience du système de santé, elle revient toujours vers le patient et la capacité du système, c'est-à-dire de l'offre de soin, de répondre à la demande au meilleur rapport coût – efficacité possible. Le sujet me semble être véritablement là. On parle beaucoup, à l'échelle de l'hôpital, d'aide à la décision médicale. On observe toutefois qu'en 2017 par exemple, 25 000 publications scientifiques traitaient de *deep learning*. Nous sommes à un moment où l'on essaie de construire les outils de demain. Une très belle *review* d'Eric Topol publiée dans *Nature* en décembre 2018 et reprenant le sujet de l'aide à la décision médicale au sein des spécialités mettait en avant un probable effet de mode, mais alertait aussi sur le fait qu'il devait y avoir un effet clinique. Or ce dernier n'est aujourd'hui pas toujours mesuré ou tout du moins publié. Les publications font souvent apparaître de belles courbes montrant qu'un algorithme peut détecter quelque chose, mais *in fine* l'efficacité et le réel avantage cliniques que ceci peut produire ne sont pas toujours corrélés. Cet aspect mérite selon moi d'être regardé, non pour contrôler l'innovation et l'investissement réalisé dans ce domaine, qu'il faut encourager, mais pour s'assurer que l'impact est réel et mesuré. Apple a par exemple déployé une montre permettant de détecter notamment les cas de fibrillation atriale, donc une pathologie cardiaque éventuelle, avec une précision assez bonne en population générale. Comme souvent, l'algorithme a été conçu pour être très précis dans la détection, quitte à détecter de faux positifs. Or ceci a un impact sur le système de soin, dans la mesure où il est aujourd'hui parfois difficile d'obtenir un rendez-vous de cardiologie. La mise sur le marché de dispositifs de ce type peut ainsi avoir un impact tout à fait pertinent à l'échelle individuelle, mais stressant pour le système de soin dans son ensemble. Je pense que ces sujets doivent être regardés avec attention.

Le NHS a publié cette semaine des recommandations relatives à la mise en place de bonnes pratiques dans la création de technologies basées sur des données de santé. Ce n'est pas de la loi narrative contraignante, mais de la *soft law*, qui part d'un principe assez sain consistant à considérer que lorsque l'on veut créer quelque chose à partir de telles données, il faut partir du besoin de l'utilisateur et voir dans quelle mesure l'utilisation des données pourrait permettre d'y répondre et quelles technologies développer, dans le respect des bonnes pratiques et des lois régissant ce type d'activité. Je pense que cette recommandation serait à examiner afin éventuellement de s'en inspirer en France pour guider le développement de ces technologies.

Les opérateurs du soin doivent par ailleurs pouvoir prendre en mains les technologies et pas simplement être en capacité de les utiliser. Il faut leur donner

les moyens de développer éventuellement eux-mêmes un certain nombre d'algorithmes et de les partager. Si l'on peut entendre la difficulté de partager des données, en revanche le partage des algorithmes, des expériences, ne devrait poser aucun problème. Ceci permettrait par exemple de ne pas avoir à développer dans un hôpital en France un algorithme pour les besoins de la recherche, alors même qu'un autre établissement a déjà mis en place un algorithme susceptible de répondre à ce besoin. Il est très important de favoriser le partage et de permettre ainsi au plus grand nombre de profiter le plus vite possible des algorithmes développés grâce à l'argent public.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Pas seulement à l'intérieur de la France d'ailleurs.

**M. Rémy Choquet.-** Ceci implique d'analyser l'applicabilité des algorithmes par rapport à la population et à ce qu'ils sont censés faire ; mais il s'agit effectivement d'un sujet important.

Pour éviter de connaître la situation rencontrée voici 25 ans dans les hôpitaux français, où chaque professionnel de santé créait son propre dossier médical, la question de l'industrialisation devra se poser. Je pense qu'elle n'est toutefois pas encore à l'ordre du jour, dans la mesure où nous en sommes encore à une phase de recherche, de développement. Il faudrait toutefois veiller à ne pas se retrouver dans cette situation de dette technologique. Les établissements de santé devront, à un moment donné, rationaliser cette mise en place d'outils. S'il existe aujourd'hui beaucoup de compétences dans nos laboratoires de recherche, d'autres sont à chercher auprès des industriels. Or ces dernières sont encore très peu utilisées par les hôpitaux, puisque l'on se situe sur un mode très collaboratif, avec des paris à cinq ou dix ans, qu'un industriel ne peut pas toujours prendre.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, M. Choquet, de rappeler qu'il existe aussi des compétences chez les industriels.

La parole est maintenant à M. Jean-Philippe Vert, chercheur associé à Mines ParisTech et au sein de Google AI. Vous êtes accompagné par Ludovic Peran, responsable des relations institutionnelles pour Google France, ce qui montre bien notre volonté d'entendre aussi la voix des équipes de recherche associées aux grands acteurs privés étrangers. Vous avez été directeur fondateur du centre de biologie computationnelle de 2006 à 2018. Vous avez en outre travaillé à l'Institut Curie sur la biologie computationnelle du cancer et été professeur au département de mathématiques de l'École normale supérieure de Paris, sur des questions telles que le développement de méthodes d'apprentissage automatique, la génomique des cancers et les apports de l'intelligence artificielle en matière de connaissance de la structure des génomes, histoire des tumeurs et assistance aux choix thérapeutiques. Vous pourrez donc aborder aussi bien des questions liées à la recherche qu'aux orientations de grands acteurs privés comme Google dans ce



domaine. Vous illustrez, par la diversité de votre parcours, la nécessité d'ouverture et d'association de toutes les forces en la matière.

**M. Jean-Philippe Vert, chercheur Google AI, chercheur associé Mines ParisTech.**- Merci beaucoup pour cette invitation et l'opportunité qui m'est ainsi donnée de présenter les sujets sur lesquels nous travaillons.

Je vais, si vous me le permettez, me présenter plus précisément, afin d'expliquer comment il est possible d'être à la fois chercheur à l'École des Mines et chez Google. J'ai étudié en France et suis mathématicien de formation. Je travaille depuis une quinzaine d'années à l'interface entre l'intelligence artificielle et la santé. Mes travaux de recherche portent plus spécifiquement sur le développement d'algorithmes, essentiellement statistiques, sur le *machine learning*, pour traiter des données de génomique dans le cadre de la recherche contre le cancer. Je poursuis ce faisant essentiellement deux objectifs : l'un relevant plutôt de la recherche fondamentale, visant à comprendre la maladie au niveau moléculaire, afin d'aboutir potentiellement à de nouveaux traitements ; l'autre, plus proche de la recherche clinique, consistant à voir comment exploiter les données génomiques pour personnaliser les traitements, les adapter à chaque individu en fonction de son patrimoine génétique et de la génétique des tumeurs, c'est-à-dire faire de la médecine dite « de précision ».

Comme l'a rappelé Cédric Villani, j'ai eu un parcours essentiellement académique et n'ai rejoint Google que l'an dernier. Aujourd'hui, j'effectue mes recherches dans le laboratoire de Google IA, ouvert à Paris en 2018. Je suis également chercheur associé à l'École des Mines. Concrètement, je continue à enseigner à l'École des Mines, à l'École normale supérieure, à encadrer des étudiants. J'interviens également à l'Institut Curie. Venant du monde académique, je puis vous dire que la recherche telle qu'elle est conçue chez Google s'apparente à la recherche académique. Le fonctionnement de la recherche dans le laboratoire Google IA au sein duquel je travaille passe notamment par des publications. Nous entretenons des liens très forts avec la communauté scientifique, par le biais d'échanges de chercheurs, de participations à des événements scientifiques. En plus des publications, un effort est également effectué pour développer des logiciels mis à la disposition de la communauté scientifique et industrielle. Beaucoup de ces logiciels sont en *open source*, ce qui signifie qu'ils peuvent être téléchargés gratuitement. Dans le cas de la génomique par exemple, le logiciel DeepVariant, développé l'an dernier par des équipes de recherche de Google aux États-Unis et ayant donné lieu à une publication dans *Nature Biotechnology*, permet de reconstituer la séquence du génome d'un individu séquencé et peut être téléchargé gratuitement sur le *web*.

J'aimerais, après cette longue introduction, profiter des quelques minutes qui me restent pour illustrer le potentiel de l'intelligence artificielle dans le domaine de la santé à travers trois exemples concrets, issus de mes travaux ou des recherches menées chez Google.

Aujourd'hui, j'observe, en tant que chercheur, que l'IA joue un rôle croissant et très important dans la recherche scientifique. Il a été largement question ce matin d'applications cliniques, diagnostiques, d'aide à la prise de décision. Mais en matière de santé, il ne faut pas oublier tout le volet lié à la recherche fondamentale, mené dans les entreprises pharmaceutiques, les laboratoires, pour trouver des cibles thérapeutiques, comprendre les mécanismes fondamentaux en jeu dans les maladies. Aujourd'hui, une grande part de cette recherche se fait de manière « *data driven* » : de nombreux laboratoires utilisent des techniques générant beaucoup de données et les outils de statistique et d'intelligence artificielle sont là pour les exploiter, leur donner du sens et permettre aux chercheurs de les comprendre. Je travaille par exemple avec des collaborateurs aux États-Unis et à l'Institut Curie sur de l'analyse de données qualifiée de génomique de cellules uniques : il existe aujourd'hui des techniques qui, face à une tumeur, peuvent mesurer à l'intérieur de chaque cellule de la tumeur (ce qui représente plusieurs centaines de milliers, voire des millions de cellules) des millions de paramètres génomique (expression des gènes, mutations, épigénomique, etc). Une telle expérience génère, vous l'imaginez, des teraoctets de données. Ce sont donc des outils d'intelligence artificielle que nous développons qui permettent de donner un sens à ces données, de comprendre par exemple dans ce cas précis l'hétérogénéité dans la tumeur, d'identifier des sous-populations de cellules susceptibles d'être responsables d'une résistance à un traitement, voire de comprendre au niveau moléculaire ce qui se passe dans ces cellules pour essayer de développer de nouveaux traitements. L'IA joue ainsi aujourd'hui un rôle fondamental dans la recherche scientifique.

Le deuxième exemple, dont il a été largement question ce matin, est l'utilisation de l'IA dans le diagnostic, notamment par l'automatisation de l'analyse d'images (radiologies, IRM, images d'anatomopathologie). De nombreux acteurs publics et privés interviennent dans ce domaine, où règne une grande compétition. Une équipe de Google a par exemple développé en 2017 un logiciel nommé LYNA (pour « *Lymph Node Assistant* »), qui se concentre, pour les patientes atteintes de cancer du sein, sur la détection de cellules cancéreuses métastatiques au niveau des ganglions lymphatiques, ce qui constitue une information cruciale pour la prise de décision médicale. Un challenge a été organisé, consistant, à partir d'images d'anatomopathologie de ganglions lymphatiques, à détecter ceux présentant des cellules cancéreuses. Google n'a pas participé à ce challenge, mais a utilisé les données pour développer un algorithme. Dans le cadre de ce challenge, les données consistaient en 270 images, à partir desquelles les équipes de Google sont parvenues à développer des algorithmes qui font aussi bien que des spécialistes humains d'anatomopathologie. Bien évidemment, ce logiciel ne prend pas de décision seul, mais peut fournir une aide utile aux médecins. Des publications récentes de Google montrent que, grâce à de tels outils, un médecin peut diviser par deux le temps passé sur une image et donc multiplier par deux le nombre de cellules qu'il détecte. La technique est là. Les données sont des données publiques

et l'on peut vraiment, avec des algorithmes à l'état de l'art, faire progresser certains domaines.

Je terminerai en évoquant l'exemple du diagnostic de la rétinopathie diabétique. Google et d'autres ont développé des algorithmes sur des jeux de données annotées, des fonds d'œil dont certains avec de la rétinopathie diabétique et d'autres non. Il s'agit là aussi d'un domaine où la performance atteinte par l'algorithme avoisine celle d'un ophtalmologue. La FDA a d'ailleurs, comme ceci a été indiqué précédemment, autorisé récemment l'utilisation de cette technique sans supervision humaine. Google dispose d'un algorithme similaire et vient de lancer en Thaïlande une grande campagne, en accord avec le ministère de la santé, pour tester cinq millions de personnes diabétiques, dans un pays qui ne compte que 1 400 ophtalmologues. Ceci illustre bien le potentiel offert par l'utilisation de ces outils dans le monde, en matière notamment de lutte contre les déserts médicaux. Google est une entreprise mondiale et doit se poser la question de savoir comment aider aux progrès de la médecine dans ces régions.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, M. Vert.

Nous en arrivons à la dernière intervention de cette matinée, qui va nous permettre de donner à nouveau la parole à Mme Bossi-Malafosse, avocate associée chez Delsol avocats, qui pourra aborder des questions telles que l'incitation au partage des données ou l'état actuel du droit et son évolution souhaitable.

**Mme Jeanne Bossi-Malafosse, avocate associée, Delsol avocats.-** Je vais tenter de proposer une synthèse de tout ce qui a déjà été dit et de rattacher ceci au droit, puisque nous sommes dans la perspective de l'examen prochain d'un projet de loi. Je pense que le fait de compléter demain le système national des données de santé avec l'ensemble des données produites par les professionnels de santé en vie réelle serait une avancée majeure. Ainsi, il ne devrait plus être possible à l'avenir de produire du soin sans produire de la donnée.

Ceci étant dit, encore faut-il mettre ce dispositif en place. Il existe sur ce point un chantier majeur d'urbanisation des systèmes d'information de l'ensemble des professionnels de santé à mener, que ce soit en ville ou à l'hôpital, pour savoir quelles données transmettre dans ce fameux SNDS et sous quelle forme. Plusieurs intervenants ont évoqué la question de la structuration des données. Faut-il des données structurées ? Faut-il aussi des données non structurées ? Il s'agit d'un problème majeur.

Ceci rejoint en outre la question, fondamentale, de l'interopérabilité : on ne pourra pas alimenter le SNDS avec ces données si l'on ne travaille pas davantage à l'interopérabilité à la fois technique et sémantique des systèmes d'information. Il est en effet essentiel d'utiliser le même langage pour décrire la situation d'un malade. Des travaux ont déjà été menés dans ce domaine, mais la

France n'est guère en avance. D'aucuns évoquaient le numéro de sécurité sociale (INS) : qu'il soit en clair ou pseudonymisé, ce numéro constitue un référentiel d'interopérabilité majeur, puisque l'on est capable dorénavant grâce à lui de certifier l'identité d'un patient et surtout de rattacher les bonnes données aux bons patients, ce qui est essentiel. Cela me ramène à mon propos de début de matinée incitant à la publication des arrêtés rendant ces référentiels opposables, non pour le plaisir de multiplier les textes, mais pour disposer juridiquement d'une situation non ambiguë sur ce point.

La gouvernance est également un élément très important. Il faut évidemment que l'État garde la main, car il s'agit d'une fonction régalienne que de garantir la santé publique. Or tout ce dont il est question ici ce matin vise précisément à garantir une meilleure santé publique pour la population. Il ne faut toutefois pas perdre de vue que le statut des organisations mises en place est majeur pour leur bon fonctionnement ultérieur. Il faut simplifier, clarifier, développer des modèles, des services numériques et s'interroger sur la question de savoir si le recours à un statut de filiale d'un GIP ou de société par actions simplifiée ne pourrait pas être plus efficace.

Le contre-pouvoir de tout ce système est la CNIL, qui depuis quarante ans accompagne tous les projets de recherche, sans les bloquer, avec toujours beaucoup de pragmatisme. Je pense par conséquent qu'il convient de renforcer les pouvoirs de la CNIL sur ce point, sans nécessairement créer de nouveau comité chargé de se prononcer sur l'éthique, alors même que ce rôle incombe déjà à la Commission nationale informatique et libertés.

J'ai enfin noté, dans les dispositions finales du projet de loi, la ratification d'un certain nombre d'ordonnances, notamment celle du 16 juin 2016 qui vient consacrer la loi Jardé sur les recherches impliquant la personne humaine. M. Vert vient d'évoquer les essais cliniques pour l'IA : il s'agit d'un aspect fondamental. Il faut savoir que, dans le cadre de la loi Jardé, a été créée une nouvelle catégorie de recherches, dites « non interventionnelles », ce qui pose beaucoup de problèmes aux acteurs aujourd'hui, puisque par définition une recherche impliquant la personne humaine ne peut pas être non interventionnelle. Ce cadre juridique est source de difficulté et pose des questions quant à la réutilisation des données à des fins de recherche ou d'intelligence artificielle. Or la ratification de cette ordonnance telle quelle rendra cette ambiguïté pérenne. Peut-être faudrait-il saisir cette occasion pour s'interroger sur cette catégorie particulière de recherches non interventionnelles, qui pourraient très bien n'être soumise qu'aux dispositions de protection des données personnelles et non pas aux comités de protection des personnes.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Merci beaucoup, Mme Bossi-Malafosse et félicitations pour votre respect du *timing*.

Nous sommes, mes chers collègues, quelque peu en retard sur l'horaire initialement prévu. Je crois toutefois que cela en valait la peine, dans la mesure où cette audition nous a permis d'évoquer une grande variété de questions et de problématiques. Il nous reste encore beaucoup de grain à moudre dans l'optique du projet de loi qui arrivera en mars en commission, puis en séance. Je sais que ma collègue Albane Gaillot, très impliquée sur ce texte et qui a suivi avec une grande attention l'ensemble des débats, est motivée et disponible pour travailler avec l'Office sur la préparation des conclusions à tirer de cette riche table ronde.

Avez-vous quelques très rapides dernières questions pour clore cette matinée ?

**Mme Albane Gaillot, députée.-** Ma question dépasse le champ de la donnée et s'adresse à Mme Bossi-Malafosse. Il a été question d'aide à la décision et au diagnostic. Si l'on tendait vers la déshumanisation du dispositif, quelle serait la responsabilité du médecin ? Quelle serait celle de l'algorithme ? Je souhaiterais bénéficier de votre regard de juriste sur ce sujet.

**Mme Jeanne Bossi-Malafosse.-** La responsabilité du médecin reste identique à ce qu'elle a toujours été. Ce débat avait eu lieu notamment à propos des outils de télémédecine : or il était apparu que ceci ne modifiait pas la responsabilité du médecin. Le fait de faire appel à un dispositif numérique ne change rien.

Un algorithme est par ailleurs, rappelons-le, le produit d'une machine nourrie par l'homme avec ses connaissances. Ceci engage par conséquent la responsabilité du fabricant.

Il s'agit de deux responsabilités différentes, qu'il va falloir accorder dès lors que l'on pourrait un jour reprocher à un algorithme d'avoir pris la mauvaise décision ou à un médecin d'avoir pris la mauvaise décision sur le fondement d'un algorithme. Comme toujours, il faut attendre que le cas se produise pour que l'on règle la question sur le plan jurisprudentiel ou judiciaire. En tout état de cause, la responsabilité du médecin n'est aujourd'hui pas modifiée par le recours aux outils numériques. Quant à l'utilisation d'outils, ceci relève de la responsabilité du fabricant. Chacun a sa propre responsabilité et il convient d'en déterminer la place exacte au cas par cas, lorsqu'un dommage survient.

**M. Cédric Villani, député, premier vice-président de l'Office.-** Prenons le cas évoqué précédemment de l'algorithme approuvé par la FDA, qui ne nécessite pas d'intervention humaine complémentaire. On peut considérer qu'il s'agit d'un outil en tant que tel, engageant de ce fait la responsabilité du fabricant, mais qu'entre aussi en ligne de compte la responsabilité du médecin qui décide de faire appel à cet outil.

**Mme Jeanne Bossi-Malafosse.-** La situation est assez similaire à celle des dispositifs médicaux, dont certains sont des objets connectés. A d'ailleurs été

publié hier le nouveau référentiel d'approbation de ces objets connectés lorsqu'ils sont considérés comme des dispositifs médicaux, sous l'autorité de la HAS. Il s'agit de certifier le bon fonctionnement d'un outil qui va ensuite être utilisé dans un cadre particulier par un professionnel de santé. Ce sont deux responsabilités différentes.

## CONCLUSION

**M. Cédric Villani, député, premier vice-président de l'Office.-** Chers amis, chers collègues, cette matinée a été placée sous le signe d'échanges très riches, dans lesquels les mots clés de souplesse, d'harmonisation, d'interopérabilité et de mise en œuvre effective se sont invités. A également été mise en exergue l'idée de rassembler tous les acteurs, publics et privés, professionnels de santé et concepteurs d'algorithmes, derrière un objectif partagé visant à l'amélioration des soins et de la qualité de vie des patients.

Nous avons vu, avec certaines comparaisons internationales notamment, que nous avons beaucoup à gagner, en particulier en termes d'efficacité, d'effectivité et qu'il était important de conserver à la CNIL le rôle de garde-fou dont elle s'acquitte avec succès, sans multiplier les outils et les filets de sécurité supplémentaires, en gardant en tête l'importance de la capacité d'action et la nécessité de pouvoir réellement transformer, aussi bien dans les aspects techniques et technologiques que juridiques et légaux, les nouvelles organisations des données en soins de santé effectifs.

Cette table ronde sera analysée en détails par l'Office dès le début du mois de mars. Ces réflexions se traduiront par des propositions formulées à l'occasion de l'examen prochain du projet de loi à l'Assemblée nationale et bien sûr, dans les années à venir, par la poursuite de ce travail qui, comme l'a souligné Olivier Clatz, montre qu'au-delà des difficultés et des obstacles il existe réellement une marge de progression importante de la part de la puissance publique sur ces sujets.

Je vous remercie toutes et tous pour votre participation et votre assiduité. Je suis sûr que nous aurons prochainement l'occasion de nous revoir pour échanger et réfléchir sur des thèmes similaires ou voisins.

*La séance est levée à 13 heures*

## Membres présents ou excusés

### Office parlementaire d'évaluation des choix scientifiques et technologiques

Réunion du jeudi 21 février 2019 à 9 heures

#### Députés

*Présents.* - Mme Valéria Faure-Muntian, M. Claude de Ganay, M. Thomas Gassilloud, M. Pierre Henriot, M. Cédric Villani, Mme Albane Gaillot, M. Cyrille Isaac-Sibille

*Excusés.* - M. Philippe Bolo, M. Christophe Bouillon, Mme Émilie Cariou, M. Jean-François Eliaou, M. Jean-Luc Fugit, M. Antoine Herth

#### Sénateurs

*Présents.* - Mme Annie Delmont-Koropoulis, Mme Véronique Guillotin, M. Gérard Longuet, Mme Angèle Prévile, M. Bruno Sido

*Excusés.* - Mme Laure Darcos, M. Jean-Marie Janssens, Mme Florence Lassarade, M. Ronan Le Gleut, M. Stéphane Piednoir