

17-38

Mars 2021

## Big Data et Droit Pénal : utilisation, compréhension et Impacts des Techniques prédictives (Drop IT)

Exemple de l'évaluation de l'incapacité totale  
de travail chez les victimes de violences

**Sous la direction de :** Thomas Lefèvre, maître de conférences – praticien hospitalier en médecine légale, HDR  
Institut de recherche interdisciplinaire sur les enjeux sociaux – Iris UMR 8156 CNRS U997 Inserm EHESS USPN.  
Hôpital J. Verdier, APHP. UFR SMBH Université Sorbonne Paris Nord

### Ont également contribué à cette recherche :

- Patrick Chariot, professeur des universités – praticien hospitalier en médecine légale. Institut de recherche interdisciplinaire sur les enjeux sociaux – Iris UMR 8156 CNRS U997 Inserm EHESS USPN. Hôpital J. Verdier, APHP. UFR SMBH Université Sorbonne Paris Nord
- Sabine Guez, postdoctorante, anthropologue, Institut de recherche interdisciplinaire sur les enjeux sociaux  
Iris UMR 8156 CNRS U997 Inserm EHESS Université Sorbonne Paris Nord
- Vincent Laugier, Tekliko SARL Singapour & Paris
- Claire Saas, maîtresse de conférences des universités en droit pénal et sciences criminelles, Université de Paris Saclay, Centre d'études et de recherche en droit de l'immatériel (CERDI)

Le présent document constitue le rapport scientifique d'une mission réalisée avec le soutien du GIP Mission de recherche Droit et Justice (convention n°217.11.23.28). Son contenu n'engage que la responsabilité de ses auteurs. Toute reproduction, même partielle est subordonnée à l'accord de la Mission.

La note est composée de cinq parties : une première partie rappelant le contexte et les objectifs du projet Drop It tel qu'il a été conçu ; une partie rendant compte de 29 entretiens réalisés auprès de magistrats du siège et du parquet, issus de plusieurs juridictions à propos des certificats médicaux avec incapacité de totale de travail et l'utilisation du numérique et de l'intelligence artificielle dans ce contexte ; une troisième expose la situation française en médecine légale, en termes d'outils numériques et d'intelligence artificielle, pour la formation, la recherche et la pratique ; une quatrième expose les principaux résultats concernant les premiers algorithmes d'aide à la décision en matière d'incapacité totale de travail ; enfin, une dernière partie présente les conclusions générales et les perspectives ouvertes par le projet.

Une partie de ces travaux a déjà fait l'objet d'une valorisation scientifique, sous forme de communication orale ou écrite, ou en contexte de formation et d'enseignement. Les résultats présentés concernant les développements techniques, à savoir, le développement d'algorithmes n'ont pas encore été soumis à la communauté scientifique, et n'a donc pas fait l'objet d'une évaluation par les pairs. Il est ainsi important de prendre les conclusions présentées pour ce qu'elles sont à ce jour : nos analyses, sur la base de l'organisation de la structuration de tout un dispositif sociotechnique de recherche, de résultats préliminaires. Nous invitons les lecteurs à considérer ces résultats et premières conclusions avec les précautions nécessaires. Les articles scientifiques correspondant sont en cours d'élaboration.

## 1 – Introduction – objectifs du projet Drop It

Nous rappelons ici les objectifs du projet tels qu'ils ont été décrits à la soumission du projet à la Mission de recherche Droit et Justice, et amendés à la demande de la Mission.

### Objectifs

L'objectif principal est le développement d'outils numériques de médecine légale contextualisée, permettant une estimation de l'incapacité totale de travail au sens pénal assise sur des critères harmonisés. Le résultat de cette étude pourrait s'inscrire dans le prolongement de la recommandations de la Haute autorité de santé (HAS) pour l'établissement d'un certificat médical initial concernant une personne victime de violences (2011), et constituerait en outre un outil utile pour échanger avec le ministère de la santé et celui de l'intérieur dans le cadre du futur comité interministériel qui sera chargé d'une réflexion sur les bonnes pratiques en matière de médecine légale (recommandation n°3 du rapport d'évaluation de l'organisation de la médecine légale en Île-de-France).

- 1) Evaluer la compréhension et l'impact possible de l'introduction de méthodes numériques quantitatives ou quali-quantitatives dans la pratique courante en matière d'évaluation de l'ITT au sens pénal, que les éléments numériques produits soit par le médecin seul ou par le médecin et un algorithme prédictif, auprès des magistrats et des OPJ, mais aussi des médecins concernés
  - a. Evaluation de la compréhension, de l'utilisation et l'utilité de l'indice numérique somato-psychique accompagnant l'ITT sur les certificats médicaux délivrés aux autorités requérantes sur leur demande ; des échelles numériques d'appréciation de la douleur, de la peur et de la perception d'une menace vitale à différents moments (au moment de faits, par la suite, au moment de l'examen) ; des échelles numériques d'appréciation de la gêne fonctionnelle chez les victimes de violence, d'appréciation de la qualité de relation victime/médecin lors de la consultation
  - b. Evaluation des représentations, de la culture numérique et des attentes des acteurs du judiciaire vis-à-vis des méthodes numériques, en particulier des algorithmes dits de médecine personnalisée ou prédictifs
  - c. Evaluation de la compréhension et de l'impact de l'introduction d'algorithmes prédictifs pour l'évaluation de l'ITT chez les victimes de violence et pour l'estimation de l'âge chez les adolescents migrants dans la pratique des acteurs du judiciaire (officier de police judiciaire, magistrats).
- 2) Participer au développement organisationnel et technique pilote d'une médecine légale contextualisée en matière d'ITT au sens pénal
  - a. Développer une application métier et en assurer sa dissémination, avec le soutien de la société française de médecine légale (SFML)
  - b. Financer une personne ressource nécessaire à l'articulation du projet entre centres, associée à l'intégration des sources, et à l'intégration de l'algorithme au logiciel métier
- 3) Objectifs secondaires et bénéfices attendus de l'étude :
  - a. Harmonisation de la discipline possible
  - b. Harmonisation des pratiques en matière d'ITT
  - c. Amorcer une base commune pour de futurs travaux i) en médecine légale, prolongeable par d'autres appels d'offre nationaux de type ANR ou PHRC et ii) côté

justice, avec l'exploitation des sources judiciaires pour les professionnels, les victimes et la recherche

## 2 – Paroles de magistrats : certificat médical, incapacité totale de travail et intelligence artificielle

### a. Méthodes : questionnaire et entretiens

Les acteurs concernés par la détermination d'une incapacité totale de travail, à commencer par les magistrats, se trouvaient au cœur du projet. Deux approches méthodologiques ont été mobilisées pour interroger les magistrats :

- Une approche consistait en la création d'un questionnaire en ligne, avec schéma randomisé pour une partie du questionnaire et visant à recueillir le maximum de réponses, même incomplètes (les informations correspondant aux objectifs principaux faisant l'objet de questions en début de sondage). Vingt-deux juridictions ont été sélectionnées, selon les critères suivants : taille (selon la typologie usuelle de classification des juridictions) et géographie, sensibilité présumée aux problématiques du numérique. Le mode de contact des magistrats suggéré par la Mission a été de solliciter procureur.e.s et président.e.s, de telle sorte que le questionnaire soit transmis par leur intermédiaire et avec leurs informations et accord aux magistrats du tribunal. Le questionnaire a été établi et revu collégialement, avec notamment l'implication de la Mission et de la Direction des affaires criminelles et des grâces (DACG). Il comportait une partie avec deux cas pratiques, où plusieurs conclusions étaient proposées aux magistrats, en termes d'orientation pénale, selon qu'un algorithme d'aide à la décision à la détermination de l'incapacité totale de travail était impliqué ou non. Les séquences de conclusions étaient randomisées.

Une seconde approche consistait en la réalisation d'entretiens semi-dirigés, réalisés par S Guez (anthropologue), basés pour une partie sur les items principaux du questionnaire en ligne, dont les cas pratiques. Deux vagues d'entretiens ont été menées : une en 2018, auprès de 19 magistrats d'Ile de France et s'intéressant en préliminaire à la compréhension et aux incompréhensions des magistrats en matière d'incapacité totale de travail et de certificats médicaux d'examens de victimes de violences ; une seconde en 2019, auprès de 11 magistrats de France (Amiens, Bobigny, Paris, Melun et Orléans) et s'intéressant directement au numérique et à l'intelligence artificielle, reprenant des éléments du questionnaire en ligne.

### b. Résultats concernant les certificats médicaux avec incapacité totale de travail

La première vague d'entretiens s'est déroulée de mars à septembre 2018 et a été menée par S Guez. Elle visait à explorer les compréhensions, incompréhensions, utilisations et attentes des magistrats en matière de certificats médicaux avec incapacité totale de travail. Ce premier travail était conçu comme un travail préliminaire au positionnement d'éventuels algorithmes d'aide à la décision en matière d'incapacité totale de travail. Dix-neuf entretiens ont été menés en Ile-de-France (magistrats du siège et du parquet de Bobigny et Paris). Tous les entretiens ont été enregistrés avec l'accord des magistrats, et ont été transcrits. Ils ont fait l'objet d'un rapport de synthèse de S Guez, « Incapacité totale de travail et certificats médicaux : paroles de magistrats ».

Les principaux points et résultats suivants ressortent du premier volet de l'enquête par entretiens :

- L'enquête qualitative réalisée auprès de magistrats en Ile-de-France au printemps-été 2018 visait à comprendre leurs usages, leurs attentes et leurs incompréhensions en matière de certificats médicaux avec incapacité totale de travail (ITT), dans les faits d'atteintes volontaires à l'intégrité des personnes et notamment les violences sexuelles. La question de l'hétérogénéité des modalités d'évaluation par les médecins de la gêne fonctionnelle résultant des faits de violence en était le fil rouge. Il s'agissait, d'abord, de cerner le poids relatif de l'ITT et du certificat dans la décision judiciaire et, ensuite, d'apprécier la perception par les magistrats de ces variations d'un certificat médical ou d'un médecin évaluateur à l'autre et de commencer d'entrevoir leurs manières de s'accommoder à cette variabilité.
- Dix-neuf entretiens semi-directifs ont été menés aux tribunaux de grande instance de Bobigny (quinze) et Paris (quatre), auprès de dix magistrats du siège (six hommes, quatre femmes) et de neuf magistrats du parquet (six femmes, trois hommes). Le groupe d'enquêtés maîtrisait des contentieux variés les exposant très régulièrement aux certificats ; il comprenait cinq magistrats dans leur premier poste et une majorité d'autres cumulant près de deux décennies au moins et jusqu'à plus de trois décennies de métier (dix magistrats dans ce cas). Leur âge moyen était de 46,7 ans. Pour la restitution du matériel inédit recueilli, le choix a été fait de privilégier la parole de ces magistrats s'exprimant sur leur pratique.
- Les magistrats interrogés établissent une distinction qui semble unanime quant à l'utilisation faite des certificats, selon qu'ils ont été établis par des médecins au sein d'une unité médico-judiciaire ou par des médecins « non légistes ». Les premiers sont considérés comme des éléments probants pour la procédure et ne sont majoritairement pas remis en cause ; les seconds sont considérés comme étant plus à même d'être questionnés.
- Si les magistrats interrogés attendent que le certificat médical et le nombre de jours d'ITT soient présentés conjointement en un même document, ils en ont des utilités et usages différents. L'ITT interviendra surtout aux « extrémités » de la procédure. Au tout début, en qualifiant les faits, elle a une fonction d'aiguillage de la procédure et un effet d'étiquetage du dossier (elle est aussi susceptible, plus tard, d'asseoir une culpabilité). Dans la dernière phase du procès pénal, l'ITT jouera un rôle déterminant en tant que critère d'appréciation de la recevabilité de la procédure devant la *Commission d'indemnisation des victimes d'infractions* (CIVI) ; elle permettra d'évaluer le préjudice et les possibles indemnisations. Concernant le cœur de la procédure, en particulier s'agissant de l'orientation pénale et de l'éventuel quantum de la peine, l'essentiel reposera sur les caractéristiques du mis en cause.
- Le certificat médical, en tant qu'il objective les conséquences psychosomatiques des violences ou des faits rapportés par le plaignant, et à condition qu'il ait été établi par un médecin légiste, est vu comme un élément de preuve majeur de leur réalité et matérialité. Les constatations médicales qui y sont consignées peuvent asseoir la crédibilité du plaignant (et leur absence au contraire le décrédibiliser). Considéré comme une pièce maîtresse – parmi d'autres éléments probatoires –, il sert à chaque phase de la procédure, nourrit l'argumentation du magistrat, et est souvent déterminant, notamment dans les violences de parent sur enfant et les violences conjugales. À ce titre, il est demandé que le certificat soit le plus complet et le plus détaillé, précis possible (la mention par le médecin d'une compatibilité des lésions décrites avec les faits rapportés sera, par exemple, appréciée si et seulement si les descriptions pour l'étayer sont circonstanciées ; « Il ne s'agit pas de dire que la victime dit la vérité mais est-ce que les constatations médicales objectivent les déclarations ? »). Une majorité de magistrats estime nécessaire une réflexion sur la formulation et les termes employés dans les certificats, en particulier les abréviations non explicitées ou les termes médicaux sans explication apportée en un langage plus accessible pour les différents lecteurs – des magistrats insistant sur le fait

qu'ils lisent aux plaignants et auteurs présumés des parties de certificats, afin d'appuyer les dires.

- La majorité des magistrats soulignent l'importance de la prise en compte du retentissement psychologique des violences, et la nécessité que celui-ci soit d'une part au mieux rapporté dans les certificats (but d'illustration et de rendre concret le retentissement évalué), et d'autre part mieux pris en compte en termes de nombre de jours d'ITT. Face à des violences sans lésion apparente, un certificat décrivant de manière circonstanciée le retentissement psychologique a une importance majeure ; sa valeur probante dépendra des détails qu'il contient. En l'absence de détails consignés par les médecins en quantité et de qualité suffisantes, les avocats de la défense plaident fréquemment l'exagération et contesteront le nombre de jours d'ITT.
- En résumé, nous avons constaté une grande déférence des magistrats à l'endroit des médecins légistes, dont la scientificité des avis et le professionnalisme leur sont d'une importance cardinale, ainsi qu'une satisfaction globale de leur part en matière de certificats médicaux et de détermination d'ITT. Cette satisfaction cohabite avec des attentes non remplies du point de vue des magistrats, au premier rang desquelles figure l'insuffisante prévisibilité du quantum de l'ITT. Les magistrats reconnaissent s'interroger fréquemment sur les différences apparentes de nombres de jours d'ITT pour deux situations de violences rapportées de manière similaires, sur la corrélation entre nombre de jours d'ITT et types de lésions ou de retentissement fonctionnel (physique ou psychologique), et enfin sur la « sévérité » de certains légistes attribuant peu de jours d'ITT au regard des situations de violences et des lésions et du retentissement fonctionnel décrits dans le certificat. Des magistrats formulent le souhait d'une harmonisation des pratiques de détermination d'ITT aujourd'hui trop souvent divergentes.
- Malgré ces questionnements, tous les magistrats interrogés ont précisé que bien qu'ils en ont le pouvoir aucun ne modifie ou revient sur le nombre de jours d'ITT établi par le médecin légiste. Cela dit, les magistrats du siège exercent leur liberté d'appréciation si d'autres éléments de preuve dans la procédure montrent que la gravité des faits est en deçà ou va au-delà de la durée d'ITT fixée. Dans ces cas rares où ils prennent leurs distances avec cette durée, ces éléments de preuve additionnels serviront à motiver leur jugement, seront pris en compte dans le quantum de la peine et auront plus tard une incidence sur la liquidation du préjudice.
- L'index somatopsychique, indicateur numérique compris entre 0 et 6 rendant compte de la part relative du somatique et du psychique dans le nombre de jours d'ITT fixé par le médecin, divise les magistrats quant à sa compréhension avant tout, et son utilité ou utilisation ensuite.

### c. Résultats concernant la perception des magistrats du numérique et de l'intelligence artificielle

Une seconde vague d'entretiens s'est déroulée de mai à octobre 2019 et a été menée par S Guez. Elle visait à explorer les représentations des magistrats en matière de numérique et d'intelligence artificielle, et esquisser ce que l'introduction d'un algorithme d'aide à la décision pourrait induire comme modification dans l'orientation pénale des magistrats étant donnés deux cas pratiques (ceux du questionnaire en ligne précédemment cité). Onze magistrats ont été interrogés. Les entretiens ont tous été enregistrés avec l'accord des magistrats. Ces entretiens semi-directifs ont été conçus dans un second temps, devant l'échec du questionnaire en ligne, et ils ont pu être transcrits. Leurs enseignements principaux ont fait l'objet d'un chapitre de l'ouvrage collectif *Sciences et sens de l'intelligence artificielle* dirigé par J Guilhem (Dalloz, octobre 2020).

Les principaux points et résultats suivants ressortent du second volet de l'enquête par entretiens :

- Les 11 magistrats interrogés se répartissaient de la façon suivante : 4 en Ile-de-France, 7 hors Ile-de-France. Sept étaient ou ont occupé des fonctions de président.e de tribunal, 4 de procureur.e ou de substitut.e. Six étaient des femmes. Deux magistrats ont fait état d'une expérience personnelle particulière du numérique et de l'IA dans le cadre de leur profession (projet d'IA dans leur juridiction, développement de logiciel ou conséquences de l'utilisation de l'IA dans un procès).
- Un certain scepticisme, sinon une inquiétude diffuse, était noté en préambule pour la plupart des magistrats interrogés, quant à l'intérêt de l'IA en droit pénal
- Interrogés sur l'utilité d'indicateurs numériques, quantitatifs dans l'évaluation du retentissement fonctionnel des violences, la majorité des magistrats considérait important et intéressant d'en disposer, en particulier en complément d'informations littérales et qualitatives. Ceux se disant le plus en faveur d'indicateurs numériques sont ceux ayant l'expérience du dommage corporel, et de l'utilisation de la nomenclature Dinthilac
- Sur l'intérêt de disposer de conclusions probabilistes, les avis étaient tranchés et homogènes : deux magistrats voyaient un intérêt réel à disposer d'une mesure d'incertitude, les 9 autres magistrats interrogés étaient de l'avis partagé qu'une probabilité permettrait d'introduire du doute, ce qui n'est pas souhaitable (besoin d'une décision tranchée, ne pas allonger le temps judiciaire...)
- Aucun magistrat ne pouvait définir ce qu'étaient la sensibilité et la spécificité d'un test, d'une mesure. Ces éléments sont déterminants et non suffisants pour avoir une idée de la fiabilité d'une mesure, ou d'un algorithme
- Sur l'utilisation d'un algorithme d'aide à la décision en matière de nombres de jours d'ITT, les avis recueillis appellent à plusieurs nuances :
  - o Un algorithme qui proposerait un nombre de jours d'ITT unique, sans information sur la variabilité probable (incertitude basée sur l'ensemble des pratiques des médecins légistes de plusieurs centres) de ce nombre de jour, était largement préféré : une mesure d'incertitude apporterait un doute, non souhaitable
  - o Néanmoins, plusieurs magistrats expliquaient que si une fourchette de nombres de jours d'ITT était proposée, ils s'en saisiraient de différentes façons : i) selon que le magistrat est du siège ou du parquet, il choisirait la valeur inférieure ou la valeur supérieure de la fourchette ; ii) selon la conclusion souhaitée, la variabilité permettrait au magistrat de choisir une valeur correspondant à sa démarche ; iii) en particulier, si le retentissement psychologique, et par exemple en cas d'index psychosomatique tirant vers une prise en compte majoritairement somatique du retentissement fonctionnel, lui semblait sous-estimé ou insuffisamment pris en compte, il pourrait utiliser des valeurs dans la fourchette supérieure.
  - o En cas de variabilité, si le médecin préconisait un nombre de jours d'ITT différent de celui présenté comme le plus probable par l'algorithme : en cas de différences faibles, les magistrats suivraient plutôt l'avis du médecin. En cas de différences importantes, ils seraient particulièrement intéressés et attentifs aux éléments qualitatifs apportés par le médecin.
  - o Ces remarques ne variaient pas selon le cas proposé : qu'il s'agisse de violences uniques par des inconnus (vol avec violence), ou de violences répétées (violences conjugales)



- La valeur de l'index somatopsychique pouvait orienter le magistrat dans sa décision de suivre plutôt la valeur de l'algorithme ou du médecin, selon que le retentissement psychologique semblait sous-estimé ou insuffisamment pris en compte
- L'utilisation d'un algorithme d'aide à la décision initialement conçu pour aider le médecin dans son évaluation, apparaissait comme également un outil potentiel d'aide à la décision indirect pour le magistrat
- A propos des caractéristiques qu'un algorithme d'aide à la décision en matière de nombre de jours d'ITT devrait prendre en compte, à une exception près, l'ensemble des magistrats interrogés considéraient :
  - Que toutes les caractéristiques concernant le médecin, le délai d'examen, le plaignant, les circonstances des violences devaient être prises en compte
  - Que les caractéristiques concernant le mis en cause ne devaient pas être prises en compte
  - L'exception mentionnée consistait à dire que les caractéristiques de la victime ne devaient pas être prises en compte
- Les réserves et les craintes à l'égard de l'IA étaient représentatives de celles exprimées usuellement, indépendamment du domaine ou de la discipline et ne semblaient donc pas spécifiques du droit pénal :
  - Crainte d'une « régression vers la moyenne », ou d'une forme d'harmonisation des pratiques des médecins légistes (souhaitée), qui pourrait davantage être une uniformisation peu différenciée et induite par une « pratique moyenne » favorisée par l'algorithme (non souhaitée)
  - Crainte d'un remplacement progressif, des médecins d'abord, puis des magistrats ensuite, pour un ensemble, sans doute majoritaire, de tâches qui nécessitent toujours plus de temps et de personnes, avec des ressources toujours insuffisantes
  - Crainte du *benchmarking* des tribunaux : qu'un tribunal soit choisi plutôt qu'un autre, connaissant les décisions prises en moyenne pour chacun (peu pertinent en matière pénale néanmoins)
  - Crainte d'un biais d'apprentissage de l'algorithme
- Il semble également que certaines craintes étaient liées à une méconnaissance du fonctionnement général d'un algorithme, par exemple :
  - Certains magistrats estiment qu'un algorithme peut estimer le retentissement somatique, pas le retentissement psychologique
- Enfin, certaines craintes ou réserves relevaient de questions centrales dépassant la seule question du recours à un algorithme, mais les réactualisant, par exemple : la tension entre approche collective, statistique, et approche individuelle, personnalisée. Il était soulevé par un magistrat qu'il n'y aurait pas de sens à comparer une personne à un groupe d'autres personnes. Cette tension est importante au sein d'un système dont un des principes repose sur l'individualisation.
- Il a été noté une opposition entre le médecin (humain) et l'algorithme (machine) à plusieurs égards :
  - Le médecin saurait pourquoi il détermine précisément tel nombre de jours, tandis que le résultat serait incertain si issu de l'IA
  - Le certificat médical était vu par certains comme objectif et scientifique, contrairement à l'IA

Plusieurs magistrats rappelaient qu'en l'espèce, la transformation numérique à l'œuvre dans leur profession était essentiellement représentée par moins de contact humain, plus de temps derrière un ordinateur, du matériel peu performant et faisant fréquemment défaut, ou peu adapté, peu ergonomique.

### 3 – Structuration des outils numériques et de leurs usages en médecine légale en France

#### a. Contexte et état des lieux du numérique en médecine légale en France

La médecine légale est, à l'échelle de l'ensemble des spécialités médicales, une spécialité de petits effectifs, tant en termes de praticiens que d'universitaires. N'étant pas une spécialité dite « d'organe », comme la cardiologie ou la pneumologie, elle n'a pas d'activité prescriptive spécifique (pas de traitement ou médicament de « médecine légale »). En tant que spécialité, elle ne dispose également que de peu de moyens financiers pour financer son développement ou sa recherche – pas d'industrie pharmaceutique intéressée à financer des essais, par exemple. Enfin, la discipline n'a pas développé jusqu'alors une culture importante de la recherche multicentrique, caractéristique considérée comme de premier plan pour juger de la portée et de l'utilité d'une recherche. D'un point de vue purement administratif et gestionnaire, il n'existe pas d'outil permettant, de façon homogène, fiable et à différentes échelles – un service ou le pays – de décrire l'activité des services.

Si la situation n'est à ce jour pas encore correctement documentée et référencée, il est admis que l'équipement des différents services de médecine légale en matière de systèmes d'information est au minimum très hétérogène, sinon peu ou très peu développé. Dans nombres de services, de tailles variées, le moindre travail de recherche sur un sujet donné nécessite de sortir les archives papier, de les parcourir individuellement et manuellement, puis de saisir des données dans un fichier constitué à ce seul usage. Plusieurs services ont recours à des solutions logicielles « sur étagère », parfois un peu modifiées ; d'autres utilisent une base de données Access ou des fichiers Excel.

Par ailleurs, les sources de données médicales ou non, individuelles ou non, se développent de plus en plus, et surtout, leur accès s'organise et se généralise, laissant espérer des recoupements jusqu'ici impossibles entre différentes informations.

Que ce soit pour des considérations médico-administratives (évaluation et pilotage de l'activité), la recherche ou la formation, il est à l'heure actuelle quasi impossible ou au prix d'efforts coûteux en temps ou en argent, d'obtenir des données particulières, comparables entre sites, et analysables.

Pour ces différentes raisons, il a été proposé de développer des outils numériques mis à disposition de la communauté, et permettant de combler au moins partiellement le manque actuel.

#### b. Structurer l'offre en outils numériques et intelligence artificielle en médecine légale

#### c. Le réseau ORFéAD

Drop It a contribué au projet ORFéAD-ML, initié en décembre 2016, et réunissant initialement 6 centres volontaires de médecine légale. ORFéAD-ML signifie Outils et réseaux pour la fédération, l'analyse et l'utilisation des données en médecine légale. Il part du principe que la recherche multicentrique en médecine légale est quasi inexistante, et peut s'expliquer par divers facteurs, notamment le peu

d'effectifs et de ressources mobilisables pour ce type de recherche, rapidement coûteuse. Le principe initial d'ORFéAD est de parier sur la possibilité de recueillir ou d'extraire des données de qualité suffisante à partir de la pratique courante des médecins légistes, à des fins de recherche, et ce en perturbant le moins possible le métier du légiste, de telle sorte que le recueil de données soit le plus acceptable possible par les professionnels y participant. S'il existe des variabilités importantes en termes de pratiques entre centres et entre médecins, tous exercent le même métier, et produisent en particulier des certificats médicaux – qui eux aussi peuvent être de forme et de fond variés. Néanmoins, il semblait plus facile de travailler à extraire de l'information de ces documents aux périmètres bien plus circonscrits que lorsque l'on s'attaque au même travail en considérant l'ensemble des spécialités médicales d'un établissement hospitalier par exemple.

En outre, le projet repose sur le fait que la majorité des centres pouvaient avoir un intérêt à partager ses certificats à des fins de recherche : en effet, comme vu précédemment, la majorité des services ne disposent a priori pas d'outils permettant ou au moins facilitant le recueil ou l'extraction de données à partir de leur activité, à des fins de recherche. Ainsi, la promesse d'ORFéAD était la suivante : en partageant leurs certificats, les services se voient offert la possibilité de récupérer leurs données structurées, prêtes à être utilisées pour la recherche, sans effort supplémentaire de leur part.

a) **Phase pilote validée** : une phase pilote s'est déroulée, grâce à l'appui du prestataire Tekliko et la poursuite du développement de la plateforme de fédération et d'analyses de données Spe3dLab (initié sur d'autres projets de collaboration de recherche avec la PME), entre fin décembre 2016 et fin janvier 2018. Au cours de cette phase pilote, nous avons cherché à nous assurer des points suivants :

- Les principes de base d'ORFéAD paraissent respectés et viables : envoi des certificats, extraction d'information utile sur la base de ces certificats, structuration possible de bases de données mono et multicentriques, accès aux analyses possibles via des comptes utilisateurs Spe3dLab pour les centres participants.
- Les centres participants sont d'accord pour continuer l'expérience, et passer à la phase d'industrialisation, c'est-à-dire, l'utilisation pérenne des moyens testés, pour la recherche en médecine légale
- Différents scénarios de financements et de fonctionnements techniques et humains ont été présentés et discutés

La phase pilote a donc été validée, et concernait alors un ensemble de 10 centres (Bondy, Garches, Créteil, Toulouse, Lille, Marseille, Boulogne sur Mer, Dijon, Tours, Nancy). Un ensemble de 15000 certificats ont pu être intégrés, selon des critères d'inclusion communiqués aux centres pour cette phase pilote (violences volontaires, avec ITT, chez des plus de 10 ans, et examinés à moins de 30 jours des faits, hors violences sexuelles).

Le circuit type des données et les traitements automatiques qui ont été validés sont les suivants :

- un médecin effectue sa consultation ; un certificat médical est produit, et existe sous forme électronique ; forme variant peu au sein d'un même service
- une fenêtre de navigateur internet permet de se connecter au compte sécurisé du médecin ou de l'agent administratif (secrétaire...) d'ORFéAD. Il s'agit d'une simple fenêtre, sans onglet, dans laquelle il suffit de glisser déposer le certificat
- le certificat est « lu automatiquement et découpé » par le serveur, qui affiche en retour à l'utilisateur ce qu'il a isolé comme données. L'utilisateur peut corriger, ou simplement valider.
- selon les nécessités ou les études en cours, la fenêtre peut afficher 3-4 autres champs, libres ou à menus déroulants, que l'utilisateur est invité à renseigner
- l'ensemble est envoyé au serveur, qui ajoute ces données à la base du centre et à la base multicentrique

- il existe à ce stade quatre types de variables possibles :
  - des variables standardisées, résultats de la saisie par l'utilisateur, des valeurs des champs additionnels de la fenêtre d'envoi des certificats (par ex : types de violences, avec une dizaine de types prédéfinis, les mêmes pour tous les centres)
  - des variables structurées directement identifiées et recopiées du certificat (par ex : le nombre de jours d'ITT, la date de la consultation, le genre de la personne examinée....)
  - des variables non structurées : ce sont des portions de texte, découpées du certificat, et conservées telles quelles pour un usage ultérieur (par ex : création de nouvelles variables sur des occurrences de mots, recherche textuelles de certificats...)
  - des variables structurées, créées automatiquement à partir du texte, sur la base de modèles préalablement entraînés sur un sous ensemble de données qui permette d'en contrôler la qualité (par ex, la variable « douleur »)
- un chercheur, avec les autorisations adéquates, peut se connecter à Spe3dLab (sur la PUD), et accéder à un ensemble de méthodes d'analyses de traitement des données auxquelles il a accès, sans installation nécessaire (les données restent sur la PUD, et les calculs s'y effectuent aussi).

Virtuellement, sans même d'autres sources de données, il existe une « infinité » de variables pouvant être créées à partir des certificats, via l'utilisation des parties textuelles et le recours à des algorithmes. A date, la base multicentrique compte une quarantaine de variables communes à tous les centres, aucune n'étant renseignée manuellement ou saisie ; toutes sont extraites d'une façon ou d'une autre des certificats.

**b) Un premier réseau de recherche en structuration** : à ce jour, ORFÉAD est le premier réseau de recherche établi et structuré dans le temps en médecine légale du vivant en France, probablement en Europe, disposant des moyens développés et mobilisant un nombre de centres aussi important et géographiquement bien répartis (ainsi qu'en termes de volume d'activité). Depuis, plusieurs autres centres ont fait part de leur souhait de rejoindre le réseau, ou sont en discussion : Amiens, Angers, Corbeil-Essonnes, Hôtel Dieu (Paris), Rouen, Versailles, Lagny et un centre algérien (ce qui n'est pas envisagé pour le moment). Ces centres sont actuellement en attente, car nous ne pouvons leur assurer une intégration rapide et pérenne au réseau, faute de moyens additionnels. Plusieurs moyens afin de pérenniser le réseau ont été envisagés ; certains ont échoué, d'autres ont été un succès.

Sur le volet technique, nous avons finalement pu opter pour un hébergement des données et des outils de fédération et d'analyses de données par un nouvel acteur : la plateforme universitaire de données de Toulouse (PUD). Les PUD sont des émanations locales de la très grande infrastructure de recherche (TGIR, investissement d'avenir) PROGEDO (CNRS et EHESS), dont la vocation première est d'organiser le recueil et l'accès aux données en sciences humaines et sociales. La PUD de Toulouse est pilotée par le directeur de la MSHS de Toulouse, avec qui nous avons par ailleurs des expériences de collaboration de recherche en santé. La PUD de Toulouse se démarque des autres PUD par sa valence « données de santé ». ORFÉAD ne nécessite pas des moyens techniques hors normes, mais une infrastructure pérenne, sécurisée et avec un service de maintenance classique assurant sa disponibilité permanente aux utilisateurs. Pour amorcer la phase d'industrialisation, le projet ORFÉAD a été considéré comme projet prioritaire de la PUD, et Tekliko comme la PUD a engagé un peu de temps humain pour la transition. Néanmoins, à ce jour, nous sommes sans ressource humaine pérenne dédiée au moins partiellement au projet (entre 0.5 et 1 ETP estimé pour le fonctionnement national ;

salaires mensuels estimés nets compris entre 2000 et 2400 euros pour un temps plein). Les aspects réglementaires, autre goulot d'étranglement, ont pour leur part pu être amorcés et sont désormais abordés avec une plus grande sérénité pour la suite du projet. En particulier : que la base de données soit légale (autorisation CNIL), et conforme au règlement général sur la protection des données (RGPD).

### c) Aspects réglementaires : DPO CNRS, circuit des données

Après plusieurs recherches, nous avons pu identifier et prendre attache avec la personne qui sera le DPO (*Data protection officer*) conformément aux exigences du RGPD pour ORFÉAD. Le projet étant initialement porté par T Lefèvre, via le laboratoire IRIS, quadri tutelle (Inserm CNRS EHESS USPN), et son directeur actuel relevant du CNRS, les règles d'attribution des DPO en cours actuellement font qu'il s'agit d'un DPO du CNRS qui sera en charge d'ORFÉAD (Emilie Masson). Plusieurs échanges ont eu lieu en 2019 avec la DPO, la PUD et les porteurs de projet. La marche à suivre est arrêtée depuis janvier 2020, et il sera demandé une conformité avec la méthodologie de référence 004 de la CNIL. Le circuit des données est identifié : le contenu des certificats sera envoyé depuis les centres, par voie informatique sécurisée, vers la PUD ; un serveur de la PUD procédera au « découpage » des données du certificat, notamment, il établira la table de correspondance entre les personnes concernées et leurs codes uniques. La table de correspondance sera hébergée à la PUD, de façon informatiquement étanche par rapport à la base de données pseudonymisée. Nous serons donc dans un cadre à finalités de recherche, avec des données identifiantes. Le choix de se placer dans un cadre plus contraignant réglementairement parlant d'emblée a été dicté par la vision à moyen et long termes ; en particulier, concernant les meilleures chances, pour d'autres projets s'appuyant sur les données ORFÉAD, de pouvoir apparier ou enrichir ces données avec d'autres sources de données – par exemple, des données hospitalières. Les évaluations d'impact et de risque concernant la vie privée des personnes concernées par la recherche seront menées par les services compétents du CNRS lors de ces démarches.

## 4 – A propos des premiers algorithmes d'aide à la décision en matière d'incapacité totale de travail

### a. Objectifs et performances attendues d'un algorithme d'aide à la décision

Un des objectifs du projet était d'étudier la faisabilité d'un algorithme d'aide à la décision en termes de nombres de jours d'incapacité totale de travail, dans l'idéal assis sur des données multicentriques et utilisables en « vie réelle », c'est-à-dire sur le temps de la consultation, en temps réel et avec le moins de saisie ou re-saisie spécifique possible de la part des médecins. Cet algorithme doit donc pouvoir être fiable mais aussi « transparent », dans le sens où si l'on sait ce qu'il fait et qu'il intervient, que sa présence ne soit pas un handicap à la pratique courante matériellement parlant.

Il est attendu de ces premiers algorithmes qu'ils :

- Rendent compte de la variabilité des pratiques inter médecins, et si possible, inter centres ; par exemple sous la forme d'un intervalle d'incertitude ou de probabilité, en nombre de jours
- Proposent, sur la base de pratiques observées dans un ou plusieurs centres, étant donné un nouveau cas (une nouvelle victime), un nombre de jours d'ITT prenant en compte les caractéristiques pertinentes de ce nouveau cas
- Proposent un nombre de jours d'ITT et un intervalle qui soient cliniquement et juridiquement pertinents
- Soient robustes et reproductibles : à même cas proposé, les résultats doivent être identiques

- Soient significativement différents d'un choix au hasard

## b. Deux approches d'élaboration de modèles « prédictifs »

Pour proposer un nombre de jours d'ITT pour un cas donné, des algorithmes peuvent être entraînés sur des données existantes, recueillies à partir de véritables cas et consultations. Dans tous les cas, l'entraînement des algorithmes respecte les standards reconnus, à savoir au minimum la séparation des données en deux sous-ensembles : l'un sert à l'entraînement de l'algorithme, l'autre à sa première validation. Des techniques de ré-échantillonnage aléatoire et de validation croisées sont utilisées. Plusieurs métriques caractérisant la qualité de l'algorithme par rapport aux données utilisées sont retenus (une seule métrique est insuffisante, toutes ayant des avantages et des défauts propres). Deux approches différentes peuvent s'envisager pour proposer un nombre de jours d'ITT au médecin en conditions de consultation :

- **Par segmentation et assignation** : On peut rechercher des groupes homogènes – au sens d'un nombre de caractéristiques prédéterminées – de situations de violences, puis essayer d'établir à quel groupe le nouveau cas appartient : on parle de segmentation et d'assignation. C'est un processus discret : on découpe en un nombre fini de groupes l'ensemble des possibilités. L'idée sous-jacente est que chaque groupe présente, à situation de violence comparable, une moyenne et une dispersion en nombre de jours d'ITT disjointe. C'est l'approche à la « Amazon » : Amazon, pour réaliser ses recommandations d'achats, découpe l'ensemble de ses clients en un peu moins de 200 000 groupes « homogènes », et assigne chaque client à un de ces groupes. Ce que les membres d'un groupe apprécient est suggéré aux membres du même groupe ne l'ayant pas acheté ou consulté. Dans le document en annexe, cela correspond à l'approche par *clustering*.
- **Par régression et extrapolation** : On peut chercher à établir un modèle qui permette de calculer directement le nombre de jours d'ITT, sur la base de caractéristiques à identifier. Nous sommes davantage dans une approche de régression/extrapolation. C'est une approche continue : toutes les valeurs d'ITT deviennent possibles, y compris, dans l'absolu, des valeurs non représentées dans les cas utilisés pour l'apprentissage. Dans le document en annexe, cela correspond aux approches (3 algorithmes différents) : GBM, *elasticnet*, *random forest*. Ces algorithmes sont des algorithmes classiques et connus en apprentissage automatique.

L'apprentissage des algorithmes (leur élaboration) peut se faire sur différentes sources, selon deux caractéristiques en particulier ici : sur des données structurées (données chiffrées, standardisées) ou non structurées (texte...), sur des données mono ou multicentriques. Ici, nous avons tiré profit des sources dont nous disposions :

- Données structurées monocentriques, dont on connaît la qualité
- Données non structurées ou semi structurées monocentriques (même centre que précédemment, pour comparaison), sur la base du réseau ORFéAD
- Données non structurées ou semi structurées multicentriques, sur la base du réseau ORFéAD

## c. La question de la qualification des données

Les données extraites automatiquement à partir des documents textuels peuvent être catégorisées comme suit : les données directement extraites et systématiquement présentes, ne nécessitant pas

d'approximation ou de traduction (par exemple : le sexe du plaignant, son âge), et les données plus ou moins complexes, extraites grâce à des algorithmes plus ou moins élaborés (traitement automatique du langage). Les premières, par définition, n'ont pas besoin d'être vérifiées, sinon qu'elles sont bien présentes ou non. Les secondes nécessitent que l'on en évalue la qualité, de telle sorte que l'on sache si elles sont susceptibles d'être utilisées pour des calculs et sont interprétables au plus près de ce qu'elles sont censées refléter. Dans notre cas, ces variables sont dites binaires ou booléennes, et se présentent sous la forme oui/non ou présence/absence : par exemple, présence ou non de symptômes dépressifs. La qualité de ce type de variables peut être évaluée par des indicateurs bien connus, reposant sur les notions de vrais positifs et négatifs, faux positifs et négatifs, comme la spécificité ou la sensibilité. Ces indicateurs renseignent sur le fait que l'algorithme, par rapport à un expert humain qui aurait lu les mêmes documents, a correctement classé la valeur de la variable ou non, et dans quelles proportions il s'est trompé. Pour ce faire, nous avons développé un outil permettant d'évaluer rigoureusement ces indicateurs, pour chacune de ces variables, selon un processus couramment utilisé lorsqu'il s'agit de classer ou de sélectionner des documents : on effectue un tirage aléatoire de 100 documents, que deux évaluateurs vont classer indépendamment l'un de l'autre. En cas de désaccord, un troisième évaluateur tranche le classement. Les résultats sont confrontés au classement de l'algorithme. Nous avons ainsi procédé à l'évaluation de la qualité de la majorité des variables générées automatiquement par des algorithmes simples. La majorité de ces variables présentaient une qualité allant de très bonne à excellente ou « parfaite », avec des sensibilités et spécificités supérieures à 0.85, et fréquemment de l'ordre de 0.90 ou 0.95.

#### d. Des algorithmes a priori peu pertinents

Sur la base des données collectées grâce à ORFéAD, nous avons décrit les populations rencontrées dans les différents centres participants (ici, 7 ont été retenus), puis utilisé ces données afin d'élaborer les premiers modèles d'aide à la décision en matière de nombre de jours d'incapacité totale de travail.

##### - Critères d'inclusions

Afin de maîtriser une partie de la grande variabilité de l'ITT, nous avons dans un premier temps restreint les populations d'étude selon les critères suivants :

- Délai entre les faits et l'examen médical : 30 jours maximum
- Age des victimes : 10 ans et plus
- Toute victime examinée pour violences volontaires, hors violences sexuelles

Nombre de jours d'ITT : de 0 à 15 jours maximum

##### - Apprentissage sur données structurées, monocentriques

Les algorithmes ont été élaborés sur la base de 63 caractéristiques possibles. Les analyses ont été menées sur un ensemble de 3120 cas. Les caractéristiques influant majoritairement sur l'ITT étaient le délai d'examen, la présence de fractures ostéoarticulaires, le type de victimes (personne placée en garde à vue, fonctionnaire de police, autre) et le médecin, dans une moindre mesure.

Les algorithmes obtenus sont peu utiles à la pratique clinique ou juridique en l'état : par exemple, l'identification de groupes homogènes ne permet pas de discriminer des intervalles de nombres de jours d'ITT. La majorité des groupes présente une variabilité en nombre de jours (donc une incertitude) couvrant la variabilité maximale de l'ensemble des possibilités, à savoir entre 0 et 15 jours, même s'ils peuvent être associés à des moyennes ou médianes différentes – mais la majorité entre 0 et 5 jours, aucune au-dessus de 8 jours.

##### - Apprentissage sur données non structurées, mono et multicentriques

Les mêmes analyses ont été menées sur les données du même centre, mais issues d'un autre mode de recueil (ORFéAD). Les analyses ont été menées sur un ensemble de 1780 cas. 42 caractéristiques étaient disponibles ; le médecin examinateur n'en faisait pas partie. Les deux ensembles de données (données structurées et non structurées), outre le fait qu'elles sont issues du même centre, n'étaient pas directement comparables. Les caractéristiques influant majoritairement sur l'ITT étaient le délai d'examen, la présence de lésions traumatiques et l'expression d'une douleur.

De façon assez inattendue, les algorithmes obtenus, en particulier par *clustering*, étaient un peu plus performants que dans le cas des données structurées, sur le même centre. Quelques groupes pouvaient être identifiés, possédant une variabilité un peu moindre. Néanmoins, ces modèles ne semblent pas beaucoup plus utilisables.

Les mêmes analyses ont été menées sur les données de 7 centres, sur la base ORFéAD. Les analyses ont été menées sur un ensemble de 5403 cas. En l'état, la répartition des cas entre les 7 centres était déséquilibrée et nécessitera rééquilibrage – ce qui augmentera par ailleurs la taille de l'échantillon. Les caractéristiques influant majoritairement sur l'ITT étaient le délai d'examen, le centre d'examen, la présence de lésions traumatiques, en particulier les fractures ostéoarticulaires et l'expression d'une douleur.

Les algorithmes obtenus, pour des raisons identiques à celles évoquées plus haut, sont en l'état peu utiles, car très peu discriminants ou précis.

Ces analyses et modèles ont également été appliquées aux mêmes données, mais après qualification (c'est-à-dire, estimation de la qualité des données extraites, et sélection de celles de qualité jugée suffisante). Les résultats ne différaient pas significativement de ceux obtenus sur des données non qualifiées.

### **Principaux résultats quantitatifs concernant la base de données ORFéAD :**

- l'étude de 42 caractéristiques extraites de 5403 documents médicaux (après application de critères d'inclusion appliqués à une base de 10 000 documents) provenant de 7 centres de médecine légale français de taille et de situation géographique variées met en évidence l'importance majeure et constante du délai d'examen, entre les faits rapportés et la consultation médicale, dans la détermination du nombre de jours d'incapacité totale de travail. Les autres caractéristiques influant en premier plan étaient le centre d'examen, la présence d'une fracture osseuse ou d'un hématome, l'expression d'une douleur ou d'une peur. De façon moindre, l'âge de la victime, la présence d'une plaie ou d'une ecchymose étaient contributifs. L'intensité de ces effets est néanmoins à quantifier.

- la qualité moyenne des variables extraites automatiquement des certificats est très bonne, et autorise une utilisation raisonnable (et raisonnée) de ces variables dans une perspective descriptive et épidémiologique des situations de violences rencontrées dans les différents centres de médecine légale français, pouvant servir de base à un observatoire des violences en France et à l'étude de leurs déterminants médico-judiciaires et psychosociaux.

- les premiers modèles « prédictifs » de l'opinion médicale en matière de nombre de jours d'incapacité totale de travail sont informatifs à plusieurs titres : les informations contenues dans les certificats, en tout cas, celles extraites en l'état, sont pour la plupart liées et contributives à l'ITT, mais ne sont pas suffisantes pour en expliquer la plus grande variabilité et donc de discriminer suffisamment finement entre deux situations en termes d'ITT. Les modèles actuels sont en l'état peu pertinents ou utiles pour envisager une utilisation pratique. Ils renseignent néanmoins sur le fait que les sources de variabilité de l'ITT sont probablement non mesurées (non mesurables ?), et que des caractérisations complémentaires, probablement au niveau du médecin examinateur, sont nécessaires.



Les algorithmes entraînés sur les données mono et multicentriques issues d'ORFéAD doivent être analysés plus précisément, et les données utilisées doivent être affinées avant de tirer des conclusions suffisamment fiables quant à ces premiers essais. Il faut également garder à l'esprit que les ITT étudiées ici sont avant tout des opinions médicales, et qu'il n'est pas ici question d'algorithmes à proprement parler « prédictifs ». Il ne saurait être question de « prédire » le nombre de jours d'incapacité totale de travail – c'est-à-dire, la durée pendant laquelle une victime éprouverait une gêne notable dans les activités courantes de la vie quotidienne, en lien avec les violences qu'elle rapporte. Que l'on parvienne ou non à obtenir des algorithmes qui rendent compte de façon utile de la variabilité de cette opinion médicale, situation par situation ou personne par personne, apportera en soi une information précieuse :

- Si l'on y parvient : on disposerait d'une analyse des déterminants de la variabilité de l'ITT, et serait en mesure de donner une fourchette de valeurs reflétant cette variabilité
- Si l'on n'y parvient pas : l'incapacité à établir les sources de variabilité de l'opinion parmi un grand nombre de caractéristiques, y compris du médecin, et qui rendrait calculable d'une manière ou d'une autre, cette variabilité, apporterait un argument important quant à la nature de l'exercice qui consiste à déterminer l'ITT. Nous serions dans le domaine de l'incalculable, et il faudrait retourner aux approches qualitatives pour mieux cerner encore les enjeux et les ressorts de cette détermination.
- Dans les deux cas, si l'on considère que l'évaluation du retentissement fonctionnel chez les victimes de violence est un élément central du traitement médico-judiciaire, il demeurera la nécessité de procéder à une véritable étude prospective, à mesures répétées et standardisées, des composantes de ce retentissement, afin de dépasser la seule opinion médicale.

## 5 – Conclusions et perspectives

Le projet Drop It a permis d'interroger des magistrats quant à leur compréhension et leur utilisation des certificats médicaux avec incapacité totale de travail ; d'avoir une idée plus précise du rôle de ces certificats dans la procédure et leur prise de décision en termes de peine pour l'auteur présumé. Ces magistrats ont également pu faire part de leur conception du numérique et de ce qu'il peut selon eux apporter, ou soustraire, à leur métier et à leur environnement de travail. Interroger les magistrats à une échelle plus large, nationale, a été avec la méthode retenue, un échec. Les entretiens fournissent des premières indications quant à la réception possible d'algorithmes d'aide à la décision qui, par ailleurs, concerneraient en premier lieu les médecins légistes, et plus ou moins indirectement, plus ou moins visiblement, les magistrats. Les médecins seront eux aussi interrogés, par voie électronique au début 2021.

Le principe de la démarche mobilisée par le projet Drop It est d'interroger tous les acteurs concernés par l'introduction de nouveaux outils numériques au sein d'une organisation, d'un processus donné, afin d'en évaluer les transformations à une échelle micro sociale. Les victimes et auteurs d'infraction, mais aussi les avocats, et les officiers de police judiciaire, devront être entendus à leur tour. Ce serait une entrée intéressante pour former et introduire l'équivalent de « patients experts » connus et reconnus dans certaines spécialités médicales, mais transposés à la médecine légale : victimes et auteurs, ayant été impliqués au cours de leur vie dans un processus médico-judiciaire, et qui seraient des ressources mobilisables pour les victimes et auteurs de violences, qui ont ou auront à passer par ce processus.

Au-delà des entretiens ou questionnaires individuels, il serait nécessaire, pour aller au bout de la démarche, de franchir deux autres étapes : d'une part, rassembler l'ensemble de ces acteurs simultanément, au sein d'une véritable simulation telle qu'on sait désormais les construire

rigoureusement, notamment en médecine et en sciences de l'éducation, afin d'estimer, interactions prises en compte, l'impact de l'introduction de l'outil numérique. Cette simulation, pour être convaincante, devrait être répétée plusieurs fois, et avec deux groupes : l'un se déroulant de manière classique, l'autre avec l'utilisation de l'outil algorithmique. D'autre part, si des algorithmes d'aide à la décision se révélaient *in vitro* potentiellement utiles en termes médico-judiciaires, alors il faudrait élaborer une expérience en situation d'utilisation réelle, sur une partie des infractions traitées par 3 juridictions : une juridiction témoin, deux juridictions différentes expérimentant l'outil.

L'autre volet important de la démarche Drop It repose sur la capacité à développer et tester rapidement, sur la base de données existantes mais de formes diverses (structurées, ou non structurées, en particulier, du texte), plus ou moins facilement accessibles, plusieurs familles d'algorithmes pressentis pour réaliser une fonction donnée, dans un usage donné. Le projet est à ce titre un succès : en l'absence de tout système d'information unique ou structuré, de référentiel commun de données, nous avons réussi à mobiliser une dizaine de centres autour du partage de documents médicaux non structurés et de présentations variées, dont nous avons su tirer et qualifier des données communes, fédérées en une base multicentrique.

Le projet a permis dès à présent de repérer et d'impliquer des intervenants importants pour l'enseignement. Il fournit également une infrastructure (ORFÉAD) et du matériel déjà exploitable. Notre regret tient à l'absence de transfert aujourd'hui, ou de mise en place d'action en formation continue, à destination des différents acteurs impliqués.

Le projet a contribué à un effort inédit de structuration en médecine légale, en particulier du vivant, rapprochant plus d'une dizaine de centres autour d'un projet commun qui a vocation à se pérenniser, de partage de données et d'analyses multicentriques. Nous avons donc à ce jour un écosystème naissant, censé favoriser et faciliter la recherche multicentrique en médecine légale. Cet écosystème nécessite des soutiens forts – ce qui n'est pas synonyme d'investissements financiers majeurs.

La structuration de ce réseau nous permet d'effectuer les premiers pas vers une meilleure connaissance des pratiques et des déterminants en matière de détermination de l'incapacité totale de travail, dans le cadre des violences volontaires, hors violences sexuelles. L'essentiel des connaissances était à ce jour réduit à des études monocentriques, en petit nombre.

Les différents entretiens réalisés, et les interactions avec les centres participants, les communications effectuées, ont renouvelé et renforcé le constat d'une harmonisation des pratiques nécessaire, par la formation et la concertation pluriprofessionnelle, en médecine légale et en particulier autour de l'évaluation du retentissement fonctionnel chez les victimes de violence.

De manière générale, l'écho rencontré autour des questions du numérique pose la question de l'opportunité de la constitution d'un groupe de travail pluriprofessionnel et pluridisciplinaire qui serait en charge de cette thématique. La méthode développée par ce projet, à perfectionner, peut se transposer à l'introduction de tout autre outil numérique dans le champ médical ou du droit pénal. L'élaboration de ce type de méthodes, puis d'expériences, est rapidement coûteuse. La mutualisation des personnes ressources ainsi que des questions, projets d'introduction d'outils numériques à l'interface de la médecine et du droit serait certainement une bonne chose. L'expérience des 15 dernières années montre à quel point les mêmes questions se posent systématiquement quand il s'agit de recourir au numérique au sein d'un organisme donné. Un tel groupe pourrait apporter son éclairage sur les initiatives en cours ou à venir, être source de recommandations et d'aide à la mise en place effective des initiatives.

La réalisation de ce projet a permis d'identifier les points aveugles et ce qu'il serait nécessaire de prolonger ou compléter sur le sujet de l'ITT et de l'IA. En particulier, nous avons identifié les éléments qui pourront faire l'objet d'une réponse à un prochain appel à projet, qui sera basée à nouveau sur une approche mixte à trois volets :

- volet technique : rendre robuste et accompagner la mise en production des outils développés en phase une (projet Drop It) ; financer un ou plusieurs masterants pour amélioration de la base existante ORFÉAD et de la qualité des variables ; participer au financement des ressources humaines

pour amorcer la phase de pérennisation ; explorer d'autres pistes algorithmiques (*deep learning...*) pour l'aide à la décision en matière d'ITT

- volet sciences humaines et sociales : compléter l'étude des représentations et de la compréhension de l'impact de l'outil algorithmique par les acteurs concernés (avocats, victimes, auteurs de violences), et évaluer leur mise en interaction par la simulation, avec l'appui du laboratoire éducation et pratiques de santé (LEPS) ; préparer une phase expérimentale

- volet valorisation par journées d'information et formation continue (avec un co-financement Iris par exemple)

Plus généralement, la méthode développée dans le cadre de Drop It sera éprouvée dans d'autres contextes ; nous ne doutons que peu de sa transférabilité.