

Gendarmes et industriels imaginent un nouveau logiciel pour prédire le crime

PAR JÉRÔME HOURDEAUX
ARTICLE PUBLIÉ LE LUNDI 25 MAI 2015

Le ministère de l'intérieur vient de dévoiler l'existence d'un algorithme prédictif, à l'œuvre depuis fin 2014, permettant de prédire les grandes tendances en matière de criminalité sur le territoire. Selon un appel à projet que Mediapart a pu consulter, un projet similaire, faisant appel à une intelligence artificielle capable de s'améliorer elle-même, est en cours de développement au sein de l'Institut Mines-Télécom, en partenariat avec le groupe industriel Safran.

Le titre du projet n'est pas encore définitif : « Horizon » ou « Anticrime ». Son but, lui, est bien fixé : « Développer un projet d'analyse et de prédiction de la criminalité ».

Alors que le projet de loi renseignement et ses fameuses « boîtes noires », censées détecter sur Internet les candidats au terrorisme, ont réveillé le spectre d'une société à la *Minority Report*, le Service central de renseignement criminel de la gendarmerie nationale (SCRC) a déjà un train d'avance. Certes, nous sommes encore loin du système de prédiction des crimes et d'arrestations préventives décrit dans la nouvelle de Philip K. Dick. Mais le projet « Anticrime » n'en demeure pas moins particulièrement ambitieux.

[[lire_aussi]]

La « proposition de projet R&D » diffusée par le laboratoire Teralab de l'Institut Mines-Télécom que Mediapart a pu consulter ne constitue qu'une première étape, une esquisse, d'un projet qui n'est pas encore réellement lancé. Elle reste révélatrice des ambitions des services de renseignement en matière d'usage de la technologie, et des nouveaux pouvoirs que va leur offrir le projet de loi renseignement.

« L'objectif du projet s'inscrit dans une démarche de renseignement criminel qui consiste à partir d'une compréhension de la criminalité à anticiper les phénomènes en vue d'une meilleure stratégie de

lutte en terme de prévention notamment », explique le descriptif détaillé. Or, selon les gendarmes, « *la criminalité* » ne peut « *être considérée comme un signal déterministe ou aléatoire, elle répond à des critères explicatifs qu'il convient d'identifier afin de pouvoir anticiper de nouvelles occurrences* ». Ainsi, si la criminalité ne doit rien au hasard et est le fruit de facteurs déterministes, il suffirait donc d'entrer un nombre de critères suffisant pour prédire ses prochaines manifestations.

Mais pour cela, il faut donc intégrer à l'algorithme « *un ensemble de variables les plus diverses afin de déterminer celles qui à différents échelons administratifs (villes, départements, régions) sont les plus significatives* ». Ces données seront d'un côté « *disponibles en sources ouvertes (Insee, météo, géographie)* » et de l'autre « *des données d'intérêt criminel* » transmises par le SCRC « *sous couvert d'une clause de confidentialité* ». Par ailleurs, « *le plan opérationnel nécessitant une rapidité d'action, des données non structurées pourront être intégrées, à savoir des extractions de blogs ou de réseaux sociaux (Facebook, Twitter...)* ».

Sur le papier, le projet est donc potentiellement colossal. « *Il n'existe pas à ce jour de projet de ce type dans le domaine de la lutte contre la criminalité qui englobe l'aspect descriptif et prédictif à des échelles de temps et d'espace différents en intégrant une telle variété de données* », annonce même le descriptif. « *En outre, un tel projet doit apparaître comme un véritable outil d'aide à la décision en matière de déploiement de ressources comme de mode d'action à envisager.* » D'autant plus que l'institut est associé pour ce projet à un partenaire privé de poids, la société Morpho, filiale électronique du groupe de défense Safran, né de la fusion en 2005 de Snecma et Sagem.

Quelles seront exactement les données injectées dans cet algorithme ? Quelles seront ses prédictions ? Comment ces dernières seront-elles utilisées ? La description du projet laisse imaginer un dispositif visant à prédire l'apparition de phénomènes criminels sur l'ensemble du territoire, afin de mieux répartir les moyens des forces de l'ordre. Mais avec quelle

précision ? S'agira-t-il juste de dire que dans tel quartier, on peut s'attendre à une recrudescence de cambriolages en été ? Ou les données, notamment issues d'Internet, permettront-elles de viser plus particulièrement certaines populations ?

Calcul stochastique et machine learning

Interrogée, la Direction générale de la gendarmerie nationale (DGGN) affirme ne pas avoir connaissance d'un projet « *Horizon/Anticrime* » mais confirme avoir « *des projets sur une analyse prédictive dans un cadre de situation préventive notamment au sein du Teralab de l'institut Mines-Télécom, en partenariat avec la société Morpho* ».

« *Le sujet consiste à effectuer des projections temporelles en 2015 sur différentes infractions (cambriolages de résidences principales et de résidences secondaires, vols de véhicules...) mais aussi spatiales sans que ne soit mentionné une seule adresse ou un seul nom* », précise la gendarmerie. Le but du projet est donc de travailler « *sur une représentation spatiotemporelle prédictive de fait afin de mieux comprendre les mécanismes de la criminalité* ».

Concernant les données qui seront utilisées, la DGGN affirme, contrairement à ce qu'indique le descriptif, ne pas travailler « *sur l'aspect prédictif incluant les réseaux sociaux* ». L'algorithme sera uniquement alimenté par « *des données quantitatives notamment via l'ONDRP [l'Observatoire national de la délinquance et des réponses pénales – ndlr] et l'Insee* ».

La question de l'analyse prédictive de la criminalité n'est d'ailleurs pas un sujet tabou pour les forces de l'ordre. Jeudi 21 mai à Cergy-Pontoise (Val-d'Oise), Bernard Cazeneuve a ainsi inauguré les nouveaux locaux du Pôle judiciaire de la gendarmerie nationale (PJGN) abritant le SCRC, l'Institut de recherche criminelle de la gendarmerie nationale (IRCGN), le Centre de lutte contre les criminalités numériques (C3N).

À cette occasion, le ministère de l'intérieur s'est livré à une entreprise de communication bien huilée. À la veille de l'inauguration officielle, la gendarmerie a invité quelques journalistes pour leur faire la démonstration d'un algorithme prédictif déjà en activité depuis la fin d'année 2014. Celui-ci intègre des données issues des faits constatés par les forces de l'ordre et des statistiques de l'Insee, pour ensuite fournir des cartes permettant d'analyser la criminalité et de prédire son évolution.

« *Nous construisons un modèle basé sur les infractions constatées entre 2008 et 2013. S'il est validé et se vérifie sur les faits de 2014, nous le projetons sur l'année 2015* », explique à **20 Minutes** Patrick Perrot, chef de la division analyse et investigation criminelle. Comme pour le projet « *Anticrime* », les résultats servent à mieux répartir des moyens de la gendarmerie sur le territoire. « *L'objectif est d'optimiser le ciblage face au développement de la criminalité pour être le plus pertinent possible dans notre surveillance* », poursuit Patrick Perrot, interrogé cette fois par **Europe 1**.



© Reuters

Impossible, à l'heure actuelle, de savoir si l'algorithme déjà en place repose sur la même technologie que celui que le Teralab s'apprête à développer. D'autant que, au sein du laboratoire, on relativise fortement la portée de cet appel à projet. Celui-ci ne serait qu'une ébauche et l'équipe n'a pour l'instant aucune idée des données qui lui seront fournies. En résumé, l'appel à projet sur-vendrait en quelque sorte à la fois son ampleur et les capacités des partenaires.

On peut toutefois s'étonner, dans ce cas, de certaines ambitions affichées et notamment une éventuelle commercialisation. « *Du point de vue de l'utilisateur final et de l'industriel partenaire, le niveau de performance prédictive atteint par les modèles et le*

format (interface graphique, outils de visualisation) des résultats produits par les outils d'analyse pourront permettre d'élaborer un cahier des charges pour un éventuel produit commercialisable », explique la fiche descriptive. Il est également précisé que « Morpho cherche à développer une offre pertinente d'analyse criminalistique sur le marché international, auprès des forces de police et de sécurité qui sont déjà ses clients ».

Par ailleurs, l'Institut Mines-Télécom est loin d'être un débutant en matière d'algorithme. La plateforme Teralab est ainsi un projet commun lancé par l'institut et le Groupe des écoles nationales d'économie et de statistique (Genes), lauréat en 2012 de l'appel à projets *big data* du Programme d'investissements d'avenir (PIA). Et le responsable du projet, Stéphan Cléménçon, est quant à lui le créateur de la chaire « Machine Learning for Big Data », un département travaillant déjà en collaboration avec des partenaires privés comme Critéo, Safran, PSA ou la BNP Paribas.

[[lire_aussi]]

Pour comprendre les enjeux de ces algorithmes prédictifs, et l'importance du projet Anticrime, il est indispensable de se pencher sur ce domaine méconnu – et spécialité du Teralab – à la frontière entre mathématique, statistique et informatique, qu'est le « machine learning ». Durant de nombreuses décennies, des pans entiers de la théorie des probabilités sont restés à un état purement théorique, principalement en raison du manque de puissance de calcul.

Ce fut notamment le cas du calcul stochastique, champ de recherche visant, pour simplifier, à étudier « des phénomènes aléatoires dépendant du temps », c'est-à-dire à calculer des probabilités dépendant de variables aléatoires et évoluant dans le temps. Mais la révolution informatique a apporté aux chercheurs les deux éléments indispensables à une mise en pratique : des machines aux puissances de calcul considérables et un stock phénoménal de données sur lesquelles travailler.

Le « machine learning » utilise le calcul stochastique mais y ajoute un élément : l'auto-apprentissage de l'algorithme. Celui-ci est conçu pour se reprogrammer, s'adapter et s'améliorer en fonction d'éventuelles nouvelles données, ou de résultats indiquant la nécessité d'un recalibrage. Plus qu'un algorithme, c'est donc une intelligence artificielle qui est ici au travail. « *Le machine learning, c'est "comment une machine peut-elle apprendre à décider toute seule ? Comment compresser, représenter et prédire de l'information à partir de données choisies pour servir d'exemples ?" Voilà tout l'enjeu du machine learning », explique l'institut Télécom ParisTech* sur son site.

Plutôt que des « prédictions » précises sur un événement en particulier, l'algorithme fournit des modèles prédictifs pouvant avoir de nombreuses applications : évaluer l'âge de décès d'une personne, prédire le parcours d'un internaute sur un site, prédire les fluctuations boursières... Le monde de l'entreprise a logiquement joué un grand rôle dans le développement du machine learning en finançant de nombreux projets.

Sans le savoir, nous sommes, en réalité, **déjà entourés d'algorithmes prédictifs**. Sur le site de Télécom ParisTech, Stéphan Cléménçon liste les domaines où le machine learning est déjà au travail : «*La reconnaissance automatique des visages en biométrie, la gestion des risques en finance, l'analyse des réseaux sociaux en marketing viral, l'amélioration de la pertinence des résultats produits par les moteurs de recherche et de recommandation, l'offre de sécurité dans les bâtiments intelligents ou encore, dans les transports, la surveillance des infrastructures et la maintenance prédictive réalisées à l'aide de systèmes embarqués...*» En matière de prévention de la criminalité, une solution très semblable au projet Anticrime, fondée sur le *big data* et les algorithmes prédictifs et **baptisée PredPol**, a déjà été déployée dans une soixantaine de villes.

« Définir un droit des algorithmes prédictifs »

Face à cette révolution scientifique permise par les progrès technologiques, la déontologie et le droit, eux, sont à la traîne. **L'article 10** de la loi du 6 juin 1978 relative à l'informatique, aux fichiers et aux libertés interdit bien toute décision « *produisant des effets juridiques* » prise « *sur le seul fondement d'un traitement automatisé de données* ». Mais cette protection semble désormais bien faible face aux progrès de la science et aux difficultés de contrôler ces algorithmes.

En fin d'année dernière, le Conseil d'État avait avancé quelques pistes de réforme dans son étude annuelle, intitulée « **Le numérique et les droits fondamentaux** ». Les sages y proposaient de « *définir un droit des algorithmes prédictifs* ». Il faut « *éviter que des systèmes présentés comme relevant de "l'aide à la décision" soient en réalité presque toujours suivis et commandent la décision, l'intervention humaine n'étant alors qu'apparente* », avertissait le rapport.

Le Conseil d'État demandait notamment un avis de la Cnil pour « *préciser l'interprétation du texte actuel, en prévoyant que l'intervention humaine ne doit pas être que formelle. Cette disposition pourrait indiquer les critères permettant de s'assurer du caractère effectif de cette décision, tels que les compétences et les qualifications de la personne qui prend la décision, la marge de manœuvre dont elle dispose dans le processus défini par son organisation et l'existence d'éléments d'information lui permettant le cas échéant de justifier de la prise d'une autre décision que celle proposée par l'algorithme* ».

Concernant l'utilisation des algorithmes dans les procédures judiciaires, l'étude allait jusqu'à proposer l'introduction du principe du contradictoire dans le traitement des données, en offrant de nouveaux droits aux citoyens. « *Lorsqu'une décision produisant des effets juridiques ou une mesure affectant de manière significative les intérêts d'une personne est en partie fondée sur un algorithme* », écrit le Conseil d'État, « *cette personne devrait bénéficier de garanties analogues à celles d'une procédure contradictoire.*

Elle doit pouvoir en effet être en mesure de faire valoir ses observations auprès de la personne qui prendra la décision, en produisant des arguments de nature le cas échéant à contrebalancer la proposition de l'algorithme ». Pour cela, la loi devrait « *imposer aux auteurs de décisions s'appuyant sur la mise en œuvre d'algorithmes une obligation de transparence sur les données personnelles utilisées par l'algorithme et le raisonnement général suivi par celui-ci. Donner à la personne faisant l'objet de la décision la possibilité de faire valoir ses observations* ».

Concernant le contrôle des algorithmes, les sages estimaient qu'il n'y a pas besoin d'aller voir à l'intérieur des « *boîtes noires* », souvent protégées par le secret professionnel ou le secret défense. « *L'ingénierie inversée* » permettrait de vérifier la nature d'un algorithme en étudiant les données entrées et en les comparant avec les résultats proposés.

« *Deux conséquences doivent en être déduites. D'une part, la Cnil doit disposer de moyens adéquats pour contrôler les algorithmes. Sur le plan juridique, l'article 44 de la loi du 6 janvier 1978, qui permet aux membres de la Commission et aux agents de contrôle "d'accéder aux programmes informatiques et aux données", lui donne déjà les prérogatives nécessaires. Sur le plan humain, il apparaît nécessaire de continuer à renforcer les moyens de la Cnil par le recrutement de spécialistes dotés de compétences adéquates. Les formations de "data scientists" créées au cours des dernières années par de nombreuses grandes écoles peuvent y pourvoir.* »



© Reuters

« *La question est de savoir si l'algorithme effectue par exemple une discrimination* », affirme Édouard Geffray, secrétaire général de la Commission nationale de l'informatique et des libertés (Cnil). Or, « *la loi n'est pas claire sur la question des*

*boîtes noires. Nous contrôlons les fichiers, mais nous n'irons pas voir l'intérieur. Sur la technologie en elle-même, la législation est encore à définir ». « Le 8 décembre dernier, à l'Unesco, le G29 a lancé une réflexion collective et adopté **une déclaration commune** visant à lancer le débat sur la construction d'un cadre déontologique. Nous considérons qu'il y a urgence. Il y a des choses fondamentales dans cette déclaration. Nous essayons d'avoir de manière globale une réflexion transversale en amont afin de construire ce qui est acceptable. »*

Concernant le projet du SCRC, le secrétaire général de la Cnil se veut prudent mais vigilant. *« Si ce projet est mené à bien, la Cnil sera forcément saisie. Je ne veux pas anticiper, mais s'il y a des données personnelles en jeu, il devra nous être soumis. Pour pouvoir exister, ce dispositif devra respecter la loi de 78. Sa finalité devra être légitime et le traitement des données devra être proportionné et adéquate. En tout cas, on regardera. »*

Finalement, la menace dystopique que ces algorithmes font planer n'est pas celle de *Minority Report*, une société totalement sécuritaire où les coupables seraient

arrêtés avant même d'avoir commis leur crime. Il s'agirait plus du régime politique décrit dans le roman *Hypérion* de Dan Simmons, un monde futuriste où les hommes auraient totalement déserté le champ politique au profit d'une gouvernance dictée par une collectivité d'intelligences artificielles, devenue indépendante et prenant ses décisions en fonction de calculs probabilistes.

Chercheuse en philosophie du droit au Fonds national de la recherche scientifique (FNRS), Antoinette Rouvroy a théorisé ce nouveau mode de « *gouvernance hors-sol* » via les algorithmes à travers le concept de « *gouvernementalité algorithmique* » (**lire son entretien**). *« La politique, c'est ce qui se glisse entre le mot et les choses », rappelle-t-elle. « Or, l'algorithme impose une réalité immanente, et qui n'est même pas une représentation du monde réel. Il impose un point de vue global, total, voire totalitaire, qui va nous mener droit dans le mur. » « Décider, ce n'est pas suivre une recommandation. C'est trancher dans l'incertitude. C'est cette incertitude qui donne sa valeur à la décision. »*

Directeur de la publication : Edwy Plenel

Directeur éditorial : François Bonnet

Le journal MEDIAPART est édité par la Société Editrice de Mediapart (SAS).

Durée de la société : quatre-vingt-dix-neuf ans à compter du 24 octobre 2007.

Capital social : 28 501,20€.

Immatriculée sous le numéro 500 631 932 RCS PARIS. Numéro de Commission paritaire des publications et agences de presse : 1214Y90071 et 1219Y90071.

Conseil d'administration : François Bonnet, Michel Broué, Gérard Cicurel, Laurent Mauduit, Edwy Plenel (Président), Marie-Hélène Smiéjan, Thierry Wilhelm. Actionnaires directs et indirects : Godefroy Beauvallet, François Bonnet, Laurent Mauduit, Edwy Plenel, Marie-Hélène Smiéjan ; Laurent Chemla, F. Vitrani ; Société Ecofinance, Société Doxa, Société des Amis de Mediapart.

Rédaction et administration : 8 passage Brulon 75012 Paris

Courriel : contact@mediapart.fr

Téléphone : + 33 (0) 1 44 68 99 08

Télécopie : + 33 (0) 1 44 68 01 90

Propriétaire, éditeur, imprimeur : la Société Editrice de Mediapart, Société par actions simplifiée au capital de 28 501,20€, immatriculée sous le numéro 500 631 932 RCS PARIS, dont le siège social est situé au 8 passage Brulon, 75012 Paris.

Abonnement : pour toute information, question ou conseil, le service abonné de Mediapart peut être contacté par courriel à l'adresse : serviceabonnement@mediapart.fr. ou par courrier à l'adresse : Service abonnés Mediapart, 4, rue Saint Hilaire 86000 Poitiers. Vous pouvez également adresser vos courriers à Société Editrice de Mediapart, 8 passage Brulon, 75012 Paris.