L'algorithme et l'ordre public

Philippe Baumard, Nadim Kobeissi

▶ **To cite this version:**

Philippe Baumard, Nadim Kobeissi. L'algorithme et l'ordre public. Archives de philosophie du droit, 2015, 58, pp. 269-288. hal-03228678

HAL Id: hal-03228678
https://cnam.hal.science/hal-03228678

Submitted on 18 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'algorithme et l'ordre public

Philippe BAUMARD[*] Nadim KOBEISSI[**]

RESUME. — Philippe Baumard et Nadim Kobeissi explorent la relation entre liberté, transparence et sécurité. Dans le contexte de l'invalidation du 6 octobre 2015 de l'accord « Safe Harbour » par la Cour de justice de l'Union européenne, et revenant sur les implications de l'affaire Volkswagen, les auteurs dénouent les liens réputés inextricables entre les notions de souveraineté numérique, libertés individuelles et publiques, et de sécurité. Les deux auteurs considèrent ainsi la perspective qu'une meilleure transparence, gérée avec rigueur, peut battre aussi bien l'efficience de marché, sans pour autant la délaisser, qu'une régulation excessive, ou l'opacité d'une économie numérique fondée sur le secret industriel et le secret des algorithmes »

MOTS-CLES. — algorithmes – cryptographie – codage – propriété intellectuelle – Al-Khawârizmî – affaire Volkswagen – *Digital Millenium Copyright Act – reverse engineering – backdoor*

Une société se reposant sur une algorithmique avancée peut-elle être une société *meilleure* ? Une société *mieux ordonnée* ? La généralisation de capacités algorithmiques centralisées (« Big Data ») et distribuées (algorithmes prédictifs embarqués dans les terminaux et véhicules) a engendré un système économique

[*] Philippe Baumard est Président du Conseil scientifique du Conseil supérieur de la formation et la recherche stratégiques (CSFRS), professeur des universités au Cnam, et au centre de recherche en gestion de l'École Polytechnique, et professeur visitant au sein des universités de Berkeley et Stanford (États-Unis). Il a notamment publié *Le vide stratégique* (CNRS Éditions, 2015, 3e édition. Il est fondateur de la start-up *Akheros*, lauréat du concours mondial innovation dans le domaine des algorithmes prédictifs pour la détection d'anomalies comportementales de sécurité.

[**] Nadim Kobeissi est chercheur en cryptographie appliquée au sein du laboratoire Prosecco de l'INRIA et au Cnam (Cedric). Il est notamment l'auteur des applications Cryptocat, miniLock et Bluenote qui ont popularisé un accès libre à des ressources cryptographiques pour le grand public. Il est l'auteur de la « Skype Open Letter », qui a réuni plus de 40 organisations, incluant *l'Electronic Frontier Foundation*, *Reporters without Borders* réclamant que Skype et Microsoft fasse preuve de transparence concernant les activités de surveillance menées par ces entreprises.

global capable de générer des rentes systématiques et distribuées, outrepassant aussi bien les structures économiques locales, sociales et les « habitus » dans lesquels ils s'insèrent. Ce phénomène diffère profondément de celui de la globalisation telle qu'elle fut vécue à la fin du XX^e^ siècle. Il ne repose pas sur la construction compétitive d'actifs géographiques auxquels sont attachés des marchés adressables, mais sur *l'insertion* ou la *substitution* d'un échange ou d'une transaction numérique en lieu et place d'un mode d'échange traditionnel. Dès lors, le pivot de la compétitivité et de la création de rente n'est plus l'infrastructure sociale encadrant la demande (le nœud de contrats et de relations humaines), mais plutôt la capacité à générer un actif de connaissance (*knowledge assets*) qui est devenu le cœur de la génération de rentes économiques. Uber en est un exemple. La numérisation de la relation entre la demande (le passager « potentiel ») et l'offre (la libre disponibilité d'un moyen de transport) permet dans un premier temps, *l'insertion* d'un marché adressable au sein de l'écosystème traditionnel (taxis, VTC), puis sa progressive substitution au fur et à mesure que l'algorithmique d'Uber prouve sa supériorité sur l'adressage antérieur du marché.

Dans ce dernier exemple, ce sont deux logiques qui s'affrontent : la compétitivité du modèle algorithmique repose sur sa capacité à créer une *information supérieure* pour le client, quel que soit le point d'adressage du marché. La compétitivité du modèle historique repose sur sa rente historique, c'est-à-dire sur sa connaissance intime et humaine du marché, le tissu de relations qu'elle a bâti avec les acteurs de son industrie, sa flotte de véhicules, les éventuels avantages qu'elle a pu obtenir, par son entregent, sur le territoire historique d'opération (accords abusifs de licences, barrières à la mobilité pour les chauffeurs, voies de taxi, aménagement des parkings d'aéroport, barrières réservées aux taxis aux arrivées, etc.)

L'enjeu de l'opérateur de l'algorithme est de « battre » l'équation de la performance du concurrent (le taxi, le VTC traditionnel) quelles que soient, – et en les intégrant – les variables explicatives et déterminantes de cette performance (temps d'attente, disponibilité des chauffeurs, prix par minute ou par kilomètre, avantages historiques, etc.). Le modèle algorithmique l'emporte, bien évidemment, car il a construit à partir de son apprentissage continuel et prédictif de la demande un outil d'ajustement du prix et du service dont l'avantage repose sur *l'asymétrie* de connaissance que l'algorithme permet d'obtenir. Mais pour ce faire, un système algorithmique doit pouvoir changer de façon dynamique aussi bien ses prémices que son équation. On réalise alors, lorsqu'un changement dans l'algorithme devient nécessaire pour maintenir cette compétitivité, que l'humain

est bien une *variable* et non un *acteur*[1]. Si un ajustement lourd tarifaire de -20 % est nécessaire, il sera unilatéralement propagé dans la base algorithmique (les « chauffeurs »), incluant très certainement la « perte de population » (les chauffeurs quittant le service) dans la modélisation de sa réception. L'algorithme sait donc diviser pour régner, ou tout au plus, sait-il appliquer le changement de programmation qui a été opéré par ces concepteurs et programmeurs. Les algorithmes ne connaissent pas la notion de « cynisme ».

Les modèles économiques reposant sur la création algorithmique d'une asymétrie de connaissance peuvent donc être destructeurs de *l'ordre public économique* local, mais dans une très large majorité, ils contribuent à son amélioration. La capacité de créer une connaissance instantanée supérieure à la connaissance émise et produite localement permet, par exemple, de soutenir la hausse inéluctable du trafic aérien mondial (gestion des routes, logiciels anticollision), la gestion optimale du marché des matières premières (nonobstant des variables géopolitiques turbulentes à faibles capacités prédictives), l'optimisation des essais cliniques sur le médicament, etc. D'un point de vue global, l'augmentation des capacités algorithmiques est considérée comme un bienfait du point de vue de la création d'un ordre public économique, qui a son tour, contribue à une plus grande résilience humaine, voire à la diffusion d'une justice sociale lorsque de tels modèles élargissent l'accessibilité d'un marché par sa baisse de prix à des populations en étant exclues (argument avancé, par exemple, par les opérateurs de location de court terme comme AirBnB).

Cette contribution de l'économie algorithmique aux ordres publics économiques locaux leur confère également une capacité à se soustraire à l'intervention réglementaire de l'autorité publique ; soit pour des buts sécuritaires, ou pour des motifs de régulation économique.

Si les modèles économiques historiques défendent leurs rentes avec leur ancrage sociétal, l'obtention d'avantages politiques (licences d'exploitation territoriales exclusives, oligopoles, parallélismes), leur « installation » culturelle, comment est-ce que les MEA (modèles économiques algorithmiques) protègent-ils la leur ? La réponse intuitive serait : « de la même façon », et tel est le leitmotiv de la plupart des opérateurs de MEA. Dans les faits, les opérateurs de MEA se trouvent systématiquement en situation antagoniste avec les déterminants de l'ordre public économique de leurs hôtes. Cela vient du fait que leur stratégie d'entrée repose sur l'insertion d'un mode d'échange *au cœur* du

[1] V. Masson, « En colère, les chauffeurs de VTC manifestent et boycottent Uber », *Le Figaro*, 14.10.2015.

modèle d'échange de leurs concurrents. La configuration de leur entrée sur les marchés est donc par nature parasitique, et crée les résistances dont nous fûmes tous témoins entre 2005 et 2015 (villes interdisant la location de court terme, pays interdisant UberPop, audition du Congrès américain de Google sur ses dispositifs d'optimisation fiscale, etc.). Dans l'écosystème local des MEA, dans le « système d'interaction » compris comme celui que décrit Raymond Boudon[2], la capture et la rétention de la rente tient uniquement à l'environnement contractuel immédiat, c'est-à-dire la possibilité pour l'opérateur d'un modèle économique algorithmique d'interdire le détournement de l'usage de sa plateforme. Il faut donc pouvoir contenir « la variable humaine » à l'intérieur de l'algorithme, et éviter à tout prix qu'elle s'en évade, ou qu'elle utilise la puissance algorithmique sans en rétribuer les concepteurs. Le point central de cette économie algorithmique est donc la *propriété intellectuelle*. Cela se traduit par des accords « commerciaux » inédits où la propriété traditionnelle est systématiquement substituée par son équivalente en contrats de licence d'usage et contrats de copyright. Nous ne sommes plus les auteurs de nos photographies personnelles sur Facebook, mais des « licenciés » de la plate-forme, non pas que Facebook cherche à accumuler un patrimoine mondial d'images de vacances, mais parce que la protection de la propriété intellectuelle de la plate-forme est la seule garante de sa rente. Pour ces entreprises majoritairement américaines, le texte légal protecteur et régulateur de leur rente économique globale est donc le *Digital Millenium Copyright Act of 1998*[3]. Dans une grande mesure, « l'ordre public économique » mondial est fortement dépendant de ce que ces opérateurs de MEA ont le droit, ou pas le droit, de faire sur leur territoire national (où sont stockées et exploitées les données, et où « résident » légalement ces algorithmes).

ORDRE PUBLIC « ECONOMIQUE » ET ORDRE PUBLIC « SECURITAIRE » ?

Le marché des opérateurs de modèles économiques algorithmiques (MEA) ne diffère pas, pour le reste de ses opérations, des marchés traditionnels. Ils sont aussi soumis à la compétition *entre MEAs*, et doivent également consacrer une partie conséquente de leurs revenus à la recherche et développement, au développement de nouveaux algorithmes, et à la protection de leur savoir-faire. Ces entreprises sont également génératrices d'emplois lorsqu'elles génèrent une demande précédemment peu adressable, ou lorsqu'elles créent *ex nihilo* de

[2] Boudon, Raymond (1987). « The Individualistic Tradition in Sociology », *in* J.C. Alexander, B. Giesen, R. Munch et N.J. Smelser (dir.), The Micro-Macro Link, Berkeley, University of California Press, p. 45-70. et R. Boudon, *La logique du social*, Paris, Hachette, 1979.

[3] http://www.copyright.gov/legislation/dmca.pdf.

nouveaux marchés (ex. : la consommation de médias numériques). Le seul problème est que le déploiement mondial de modèles économiques algorithmiques peut engendrer des antagonismes locaux, détruire des bassins d'emplois, déstabiliser des filières économiques, lorsque l'écart de compétitivité (*l'asymétrie de connaissance*) n'est pas mathématiquement rattrapable par l'acteur historique. Lorsque dans l'équation, la production d'une variété requise (un large choix de produits) a pour conséquence l'impossibilité économique de maintenir une activité, l'acteur historique peut soit réduire drastiquement cette variété (modèle utilisé par les grands distributeurs) en offrant une ligne de produits d'entrée de gamme, mais à marges satisfaisantes, soit abandonner le secteur[4].

« L'ordre public sécuritaire », – que l'on pourrait définir comme la protection du public, les normes de sécurité, les normes de pollution, la possibilité qu'une plate-forme soit détournée pour atteindre à la souveraineté nationale d'un pays, la destruction du bien être public par une concurrence déloyale, le respect des réglementations sur la vie privée – est donc, dans son ensemble, plutôt *une contrainte* pour les MEAs, qu'un objet de tout autre intérêt. Il est une contrainte car il empêche la standardisation logicielle. Les exceptions coûtent cher dans un code, et elles coûtent cher en *mitigation* lorsqu'elles n'ont pas été respectées. La performance économique de la rente des MEAs est directement dépendante des effets d'échelle et d'éventails que l'algorithme peut produire. C'est justement parce que ces modèles économiques algorithmiques peuvent s'absoudre des « contextes » locaux, que leur avantage de coût absolu est si important.

Le second problème pour les opérateurs de MEAs, aussi bien avec l'ordre public que la liberté publique, réside dans leurs implications pour leur propriété intellectuelle. La plupart de ces modèles économiques algorithmiques sont *embarqués*, c'est-à-dire qu'ils sont encapsulés dans un terminal, dans les composants d'une voiture, dans un système de contrôle industriel, dans une infrastructure critique, etc. Cette « encapsulation » pèse lourdement sur la stratégie de protection de la rente (propriété industrielle) : les opérateurs de modèles économiques algorithmiques cherchent à protéger leur secret industriel, les variables qu'ils recueillent et analysent, les formalisations de leurs algorithmes, si bien que l'algorithme, tout autant que le système qui le contient (capteurs, senseurs, « écoute de l'environnement »), ne font généralement qu'un.

[4] C'est ce qui se passa sur des marchés physiques (comme la distribution de produits blancs ou gris, électronique de grande consommation), aussi bien que des marchés numériques (les opérateurs de télécommunications).

Du point de vue de l'industriel du MEA, plus tôt est réglé le problème de la conformité réglementaire, plus solide est la protection de la rente. Cela se traduit dans une contradiction fondamentale avec le processus de certification et ce qui se passe après :

- Il est dans l'intérêt de la vérification et de la fiabilité d'un code logiciel qu'il puisse être vérifiable à tout moment, c'est-à-dire que son code source soit ouvert pour son libre examen, par l'autorité de certification elle-même.

- Mais dès que cette vérification est effectuée par le régulateur, les industriels ferment ces codes à toute inspection et à tout contrôle ultérieur, au motif de la protection de leur rente industrielle, ou pour tout autre motif de confidentialité stratégique.

Cela se traduit par une véritable « chaîne du froid » du point de vue la certification et de l'inspection : il est dans l'intérêt primordial pour l'industriel des MEAs que son code soit certifié le plus tôt possible dans la chaîne de valeurs (dès la production), et qu'il reste solidement protégé, obfusqué, chiffré, obscurci et/ou inaccessible jusqu'à sa livraison et son *usage* par le consommateur. C'est pour cette raison que les accords de propriété intellectuelle associés à l'usage de ces algorithmes s'efforcent à tout prix que la notion de propriété, qui en permet l'accès et la libre disposition, en soit exclue.

« Ordre public économique » et « sécurité publique » n'ont jamais été aussi présents, et autant considérés avec appréhension et mysticisme, que dans le débat qui oppose aujourd'hui les tenants d'une vision sécuritaire de la vie en société, et ceux d'une défense des libertés fondamentales : droit d'expression, droit de détermination, souveraineté morale et juridique… La société numérique n'est pas une société du chiffre ou du nombre, mais une société profondément panoptique. Elle oppose ainsi d'un côté ceux qui ont la capacité synoptique de voir, non pas « le grand nombre » propre à la logique des masses d'Elias Canetti, mais chacun *à la loupe*, « par le grand nombre », et ceux, de l'autre côté, qui sont « vus sans voir », « lus sans lire », « expliqués sans pouvoir expliquer ».

Cette capacité asymétrique de « voir sans être vu » diffère du *Panopticon* de Bentham dans le fait qu'elle ne repose pas sur l'ubiquité du déploiement ou sur l'incertitude d'être surveillé créant un effet de potentialité[5], mais sur l'ubiquité et le faible coût marginal des capacités algorithmiques d'apprentissage. La capacité

[5] *Cf.* à ce propos Michel Foucault, *Surveiller et punir*, Paris, Gallimard, 1975. Jeremy Bentham, *Panopticon or the inspection-house*, 1791 et pour une analyse de la relation entre économie numérique et dispositifs néopanoptiques, Ph. Baumard, *Stratégie et surveillance des environnements concurrentiels*, Paris, Masson, 1991.

à produire des modèles économétriques précis n'est pas nouvelle. Elle régit de nombreux arbitrages des sociétés contemporaines, dont en premier lieu celui de la mutualisation des risques assurantiels. La différence réside à la fois dans la possibilité de conduire ces apprentissages sur de plus grands nombres (le « Big Data »), améliorant ainsi leurs capacités prédictives, et d'utiliser, dans l'instant, et de manière distribuée, les résultats de cet apprentissage (la *connectivité*). L'apprentissage réalisé sur un très grand nombre permet de créer un effet massif d'externalité, c'est-à-dire de bénéficier de modèles probabilistes affinés par un apprentissage réalisé sur des centaines de millions d'individus. Cette granularité plus fine permet de produire une connaissance qui n'est pas atteignable par le niveau local où elle s'exerce, c'est-à-dire que l'extraction de la connaissance locale (ex. : comportement d'une personne, métadonnées, etc.), son traitement par des modèles d'apprentissage bénéficiant de l'ubiquité de situations permet de créer de façon continue une connaissance meilleure que l'agent local. En d'autres termes, on peut produire à un point distant une *simulation* en information pure et parfaite, d'une situation *réelle* dont l'information est imparfaite et incomplète.

Il est donc *à ce stade de l'état de l'art* impossible d'effectuer le chemin inverse : l'agent économique local ne peut pas, à partir des données dont il dispose, reconstruire une connaissance pure et parfaite de sa propre situation. Cette asymétrie générique de l'économie numérique est à la base de son modèle de rente : c'est le différentiel de connaissance, l'asymétrie, qui crée la rente immédiate, dont la valeur est justement exacerbée par son caractère périssable. Avec le mécanisme connu des économistes industriels, la conjonction de l'asymétrie et d'un périmètre de pertinence court en géographie et périssable (doublement fini), chaque situation décentralisée ainsi créée correspond à un oligopole, non pas fondé sur des matières premières, mais sur la maîtrise à distance d'un système cognitif[6].

Derrière une idéologie du progrès technique, la logique de contrôle oligopolistique des marchés reste absolument la même, sauf qu'elle ne repose pas sur des entrées expresses, des limitations directes de la concurrence par les milieux d'affaires ou des collusions protectionnistes, mais sur l'interdiction d'examen du code logiciel qui régit *de facto* le comportement du marché.

Cette asymétrie s'exprime dans tous les systèmes numériques sans exception, pourvu qu'ils soient dotés d'une capacité d'apprentissage synoptique (c'est-à-dire

6 *Cf.* Ph. Baumard, « Les contextes cognitifs de la stratégie », *in* F. Tannery, A.C. Martinet, T. Hafsi et J.-P. Denis (Eds.), *Encyclopédie de la stratégie*, Paris, Economica, 2014.

capables d'une capacité d'enrichissement de la connaissance locale supérieure à ce que ce système local peut produire lui-même). Elle crée de fait l'existence d'une connaissance *extraite* qui est supérieure à la connaissance *endogène* de l'acteur, c'est-à-dire, du point de vue de l'ordre public, que la connaissance sociétale ainsi produite (par exemple, prédiction des comportements individuels, prédiction des risques de sécurité) dépasse la capacité décisionnelle individuelle, et parfois collective, du sous-ensemble local concerné. Ce « sous-ensemble » peut correspondre à la conduite d'un véhicule automatique (Tesla, Google Cars), à l'optimisation tarifaire d'un service fondé sur des observations d'usage, de flux, de trafic de véhicules et du comportement de la concurrence (ex. : Uber), à la gestion de droits d'accès à des sites internet pour la protection des mineurs, comme ce fut le cas récemment en Corée du Sud[7].

Cette asymétrie est générée par une capacité algorithmique. La capacité algorithmique est la possibilité de produire à partir d'un ensemble de données une fonction calculable permettant de comprendre, caractériser, expliquer ou prédire l'état courant ou futur des données capturées.

DE QUELLE MATIERE EST FAITE « LA LOYAUTE ALGORITHMIQUE » ?

Un algorithme n'est en soi qu'une suite d'opérations proposant un moyen prouvable de résoudre un problème. Comme dans l'ouvrage d'Al-Khawârizmî qui en institua les principes, *L'abrégé du calcul par la restauration et la comparaison*[8], écrit en 813 et 833, les algorithmes ont pour objet de proposer des méthodes « non opposables » permettant de résoudre de façon non réfutable des problèmes communs. L'algèbre d'Al-Khawârizmî est une commande du calife Al-Ma'mûn qui poursuit également cet objectif. Il est composé d'exemples concrets pour mesurer des arpents de terre, partager équitablement des récoltes, introduisant les équations du premier et second degré pour ce faire. On peut, en ce sens, attribuer une grande paternité de l'algorithmique moderne à Al-Khawârizmî, en ce que l'ancrage de ses démonstrations sont des résolutions de problèmes de société, de gestion courante, à la différence des équations diophantiennes, du mathématicien grec Diophante d'Alexandrie, à qui l'on peut sans doute attribuer une paternité plus ancienne à l'algèbre pour son étude des équations à variables

7 Collin Anderson, John Scott-Railton, Masashi Crete-Nishihata, « Are the Kids Alright? Digital Risks to Minors from South Korea's Smart Sheriff Application », *CitizenLab*, 20 septembre 2015. https://citizenlab.org/2015/09/digital-risks-south-korea-smart-sheriff/.

8 *L'Abrégé du calcul par la restauration et la comparaison* (الجبر حساب في المختصر الكتاب والمقابلة).

rationnelles. L'algorithme propose ainsi une intermédiation sociale irréfutable entre un problème et sa solution, en ce que tout un chacun peut effectuer indifféremment les deux chemins possibles : du problème vers la solution, et de la solution vers le problème (comprendre et connaître la méthode irréfutable par laquelle le problème fut réglé).

Le principe d'une bonne performance algorithmique réside donc dans la lisibilité et la capacité de compréhension de chacune de ses étapes, de chacun de ses calculs élémentaires suffisamment basiques pour qu'il puisse être reproduit avec les mêmes résultats par celui qui l'émet, celui qui le reçoit, celui qui en fait usage. Dès lors, l'algorithme n'est pas la boîte noire néoclassique de l'économie contemporaine, mais plutôt son *antimatière* : par l'origine même de sa fondation, l'algorithme, dans la transparence de sa réfutation comme de sa preuve, devrait être un garant de stabilité et d'équité sociale, c'est-à-dire *d'ordre public*.

Le problème posé par l'algorithmique contemporaine n'est donc pas seulement celui de sa possible incomplétude, de son secret, de sa non-vérifiabilité, mais aussi une question économique : même si un algorithme est réputé vérifiable, ou que *sa vérification a été effectuée*, quels sont le coût et l'accessibilité de cette vérification par des structures de marché, par le droit, ou par son usager ?

Lors de sa fondation, la science algorithmique répondait à la fois au défi d'expliquer le monde, ce que la Grèce antique appelait *l'épistémè*, autant qu'à réguler l'intelligence sociale, *la phronesis*. Le calife Al-Ma'mûn commandite un traité d'algèbre parce qu'il a une utilité sociale, parce que l'algorithme est un régulateur des relations marchandes et un pacificateur de la société. Il n'est pas en cela différent du code de Hammurabi : un système symbolique émetteur de règles et descripteur des étapes logiques de résolution de problèmes courants. L'algorithme est à la fois le *jeu*, la *règle du jeu*[9] et le garant constitutif de l'impartialité de ses résultats.

L'affaire Volkswagen est dans ce domaine riche d'enseignements. Les logiciels embarqués de Volkswagen se mettaient en œuvre pendant l'inspection, substituant une donnée produite par un algorithme à la mesure réelle devant être

9 *Cf.* André-Jean Arnaud, « Du jeu fini au jeu ouvert. Réflexions additionnelles sur le Droit post-moderne », *Droit et société*, 1991, 17(1), p. 39-55.

produite par les instruments de mesure[10]. Pour chacun des 11 millions de véhicules, la société « avait installé un logiciel dans le module de contrôle électronique (ECM) de ces véhicules qui détectait la présence d'un instrument de test de contrôle d'émissions EPA. L'EPA a dénommé ce dispositif le "Switch". Le Switch déterminait si le véhicule était testé à partir de plusieurs variables telles que la position du volant, la vitesse du véhicule, la durée d'opération du moteur, et la pression barométrique » [...] Dès lors que ces variables indiquaient, par calcul algorithmique, la possibilité qu'un test soit en cours, le logiciel « activait un programme spécifique de calibration de route, qui réduisait l'efficacité du système de contrôle d'émission (et plus spécifiquement la réduction catalytique sélective ou le piège à NOx) ».[11] Dans le cas Volkswagen, deux algorithmes sont à l'œuvre. Un premier algorithme permet de déterminer par un simple calcul de corrélation de variables environnementales si le véhicule est sur un banc de test ; ce qui déclenche le lancement d'un programme contenant un algorithme visant à modifier les résultats du test. Dès que le véhicule retournait sur la route, le programme de « Calibration » s'auto-désenclenchait, redonnant au véhicule sa pleine puissance et 35 fois le niveau de pollution au NOx accepté par la législation américaine.

Comment Volkswagen a-t-elle pu échapper aussi longtemps à la vigilance du régulateur, avec un écart de la donnée réelle à la donnée simulée aussi important ? La réponse réside dans la protection accordée par le *Digital Millenium Copyright Act*[12] de 1998. Le DCMA interdit à quiconque d'outrepasser les « mesures de protection technologiques » protégeant du code faisant l'objet d'une propriété intellectuelle. La Bibliothèque du Congrès peut cependant, dans des exceptions rares, accorder un tel droit de regard pour des chercheurs travaillant à l'analyse de vulnérabilités. La Classe 25 de ces revendications au titre du DCMA permet le *reverse engineering*, c'est-à-dire la possibilité de briser les protections d'un code (chiffré ou « obfusqué ») pour en retrouver les étapes logiques : en d'autres termes, permettre « de faire le chemin inverse » depuis le résultat final jusqu'aux prémices et aux primitifs du résultat obtenu. Mais lorsqu'une telle demande fut formulée, l'Alliance des Manufacturiers Automobiles engagea ses meilleurs avocats pour en rejeter le pourvoi, invoquant la création « de menaces sérieuses à la sécurité des passagers »[13]. L'agence de

[10] Jim Dwyer, « Volkswagen's Diesel Fraud Makes Critic of Secret Code a Prophet » *The New York Times,* 22 sept. 2015.

[11] Extrait, page 3, de la « notice of violation », adressée à Stuart Johnson, par l'EPA le 18 septembre 2015 : http://www3.epa.gov/otaq/cert/documents/vw-nov-caa-09-18-15.pdf.

[12] http://www.copyright.gov/reports/studies/dmca/dmca_executive.html

[13] Short Comment Regarding a Proposed Exemption Under 17 U.S.C. 1201 (Proposed Class #25) submitted on behalf of The Alliance of Automobile Manufacturers ("Auto Alliance"), the leading advocacy group for the auto industry. Auto Alliance represents

protection de l'environnement (EPA) s'opposa elle-même à cette demande, redoutant que la diffusion d'une telle information puisse aider de potentiels « hackers »[14]. La crainte émise par l'EPA est qu'un accès au code puisse permettre la « customisation », c'est-à-dire donner la possibilité à des usagers de diminuer le contrôle d'émission sur leurs véhicules pour gagner en performance, c'est-à-dire faire exactement ce que Volkswagen a fait à très large échelle.

Le problème posé par l'affaire Volkswagen est bien celui de la loyauté algorithmique, et plus précisément, l'absence d'un fondement juridique permettant à un tiers de s'assurer que l'ensemble des opérations élémentaires de l'algorithme soient connues, finies, complètes, et aboutissent bien à un résultat vérifiable. Ce cas renvoie à la problématique générale de la relation entre le logiciel et les régulateurs dans de nombreux domaines : l'existence de *backdoors* dans les logiciels de chiffrement, la sécurité des messages électroniques, le caractère propriétaire et non vérifiable de logiciels implantés dans des infrastructures vitales (énergie, transports, nucléaire). La justification de l'opacité algorithmique repose généralement sur l'argument économique (rente, valorisation de la propriété intellectuelle), tandis que l'accès à des logiciels ou de l'électronique repose sur des arguments de sécurité publique (comme ce fut le cas pour Volkswagen de la part de l'industrie automobile, et du régulateur lui-même !).

Reprenons, étape par étape, le déploiement de la falsification des tests par Volkswagen, et essayons d'en tirer des enseignements en termes de concepts fondateurs d'une « loyauté algorithmique » :

- Le premier algorithme développé par Volkswagen est celui qui détermine si un véhicule est sur un banc de test. L'objectif est ici d'échapper à la détection grâce à l'intelligence artificielle, c'est-à-dire de conférer à un système ou une machine (ici, un véhicule) une capacité prédictive permanente de son environnement immédiat : une conscience situationnelle. Dans le cas de Volkswagen, l'algorithme analyse le comportement du véhicule, et par recoupement, en déduit la position immobile sur un banc de test (ex. : pas de corrélation entre les mouvements du volant et les changements de régime du moteur ; le changement de direction du volant n'engendre pas de résistance des

77% of all car and light truck sales in the United States, including the BMW Group, FCA US LLC, Ford Motor Company, General Motors Company, Jaguar Land Rover, Mazda, Mercedes- Benz USA, Mitsubishi Motors, Porsche, Toyota, Volkswagen Group of America and Volvo Cars North America.
http://copyright.gov/1201/2015/comments-032715/class%2025/Auto_Alliance_Class25_1201_2014.pdf

[14] Alex Davies, « The EPA Opposes Rules That Could've Exposed VW's Cheating », *Wired*, 18 sept. 2015.

roues sur l'asphalte, et ne produit pas de surrégime du moteur), le temps de fonctionnement du moteur dans une telle configuration, la pression barométrique, etc.

- L'algorithme agit ensuite comme déclencheur *uniquement* quand celui-ci produit une prédiction de période de test pour les émissions de pollution, ajustant le régime du moteur afin de tromper les équipements de test.

- Les concepteurs de la plate-forme logicielle embarquée se sont assurés que leur logiciel soit inaccessible, invisible à la fois à l'équipement de test et au propriétaire du véhicule. Dès que le test est terminé, le logiciel retourne dans un état de sommeil, et rétablit les réglages permettant au véhicule d'obtenir la bonne performance qui fait la réputation de la marque, en émettant 35 fois le niveau légal de NOx dans l'atmosphère.

Est-ce que cela semble familier ? C'est exactement la même logique qui est utilisée pour les communications chiffrées d'un téléphone portable. Ce qui est intéressant dans le cas Volkswagen est l'utilisation d'un algorithme prédictif suffisamment adaptatif pour pouvoir être déployé à très grande échelle. Dans le cas d'un *backdoor* installé sur un téléphone portable :

- Un algorithme embarqué estime la probabilité d'une communication présentant un intérêt potentiel pour l'autorité surveillante. La législation américaine sur les « métadonnées » peut permettre la programmation de tels algorithmes. Les métadonnées comprennent la localisation géographique d'un appel, celle de son propriétaire, l'information contenue dans l'abonnement téléphonique, l'identité numérique du terminal ayant envoyé cette information (IMEI), le moment de la journée (heure, minute, seconde) à laquelle est passée la communication, les mêmes informations concernant l'interlocuteur, et tous les interlocuteurs d'un point de vue historique, la longueur de cette communication, son intensité en données ; mais également l'historique comportemental du terminal, incluant le taux d'usage des réseaux sociaux, l'identité des « amis » sur ces réseaux sociaux, le nombre et la nature des téléchargements, les détails du point d'accès (qui n'est pas de la responsabilité du fournisseur d'accès, mais du fabricant), *à l'exception* du contenu des communications (voix et données). L'existence de tels algorithmes, sans le consentement informé de l'usager, est rendue possible justement parce que le régulateur se satisfait d'une certification de conformité très en amont d'un code dont les fonctionnalités réelles sont protégées par la propriété intellectuelle.

- Il est donc dans l'intérêt de l'industriel d'élargir autant qu'il puisse le faire, – en influençant le régulateur –, la définition des données qu'il peut légalement capturer, ou d'en maintenir l'ambiguïté attributive, lorsqu'elle peut lui permettre d'inclure des données qu'une définition opportunément vague n'aurait pas anticipées. Par exemple, aux États-Unis, le code 3127 n'incluait pas les données

contextuelles aux communications[15], et se limitait à l'identification de l'appelant et de l'appelé : « *captures the incoming electronic or other impulses which identify the originating number of an instrument or device from which a wire or electronic communication was transmitted.* ». Cette définition fut élargie en 2013, avant d'être portée par *l'American Civil Liberties Union* (ACLU) devant la cour d'appel du second circuit de justice américain[16]. La différence introduite par la nouvelle législation américaine a été d'élargir, sur les fondements de la jurisprudence, la notion de « *any device* », qui peut se comprendre d'un point de vue exclusif (seulement le terminal), ou inclusif (le terminal et toutes les applications embarquées par ce terminal : consultation de comptes bancaires, réseaux sociaux, SMS, chats, application « santé », etc.). Cette extension a été rendue possible par extension du *Foreign Intelligence Service Act* (FSIC).

Tous les pays du G8 ont peu ou prou des dispositions similaires, la plupart dans des mesures moindres que celle des États-Unis[17]. Si la donnée nominative n'est pas concédée par l'utilisateur (par exemple dans le cas d'un utilisateur utilisant un pseudonyme, une connexion anonymisée de type Tor, un DNS chiffré et une clé de chiffrement pour ses communications), il est toujours possible de « reconstruire » son identité non pas en déchiffrant ses communications, mais par la corrélation entre sa position géographique, ses amis sur Facebook, le fait que ce terminal dont l'identité unique (EMEI) est considérée une « métadonnée » a été appelé par une personne qui elle n'a pas pris de précaution de chiffrement ou de masquage par pseudonyme.

Même un utilisateur qui entrerait dans un magasin aléatoire, ayant fait attention à se rendre dans un magasin dans lequel il n'a jamais mis les pieds, payant en liquide une carte téléphonique, sans laisser de nom, peut difficilement s'extraire de la « nasse » de corrélations continuelles qui l'associent à des lieux, des contacts, des « patterns » et des modèles prédictifs. Acheter un nouveau terminal et une nouvelle carte SIM à chaque déplacement est, dans ce domaine, d'une quasi-inutilité. Pour défaire une telle puissance algorithmique, il faudrait

15 *Cf.* https://www.law.cornell.edu/uscode/text/18/3127.

16 ACLU vs. Clapper. United States Court of Appeals for the Second Circuit. Argued: September 2, 2014 Decided: May 7, 2015. Docket No. 14-42-cv, Case 14-42, Document 168-1, 05/07/2015, 1503586, 97 pages. http://pdfserver.amlaw.com/nlj/NSA_ca2_20150507.pdf.

17 Pour un exemple des métadonnées recueillies sur différents terminaux, *The Guardian* a mis à la disposition du public une page d'information expliquant leur nature et leur utilisation : http://www.theguardian.com/technology/interactive/2013/jun/12/what-is-metadata-nsa-surveillance#meta=0000000.

qu'un utilisateur, ou un groupe d'utilisateurs, puisse produire une factorisation supérieure à celle du système de surveillance, c'est-à-dire forcer la technologie prédictive à n'obtenir aucun résultat probabiliste supérieur à l'équiprobabilité (50/50), malgré la forte disponibilité d'informations certaines (lieux, tour d'émission la plus proche, lieu connu le plus récent, lieu de réapparition le plus récent, etc.). Le degré de liberté, c'est-à-dire le nombre de possibles que devrait produire un groupe d'individus (qui par exemple auraient tous choisi de s'appeler « John Doe », et s'échangeant toute communication en utilisant le même pseudonyme « johndoe ») devrait atteindre un nombre considérable afin de battre de tels algorithmes prédictifs.

- L'algorithme lui-même peut être utilisé, en mettant à profit les métadonnées présentées en exemple, pour n'arrêter le chiffrement *que lorsque la situation est adéquate*, c'est-à-dire lorsque le modèle prédictif indique que vous êtes probablement en « situation d'intérêt » (salles de conseil d'administration, ou, *justement*, dans un lieu très improbable ou incongru connaissant vos habitudes d'usage). Il peut ainsi être programmé pour arrêter le processus de chiffrement, ou le rendre vulnérable à un accès par un tiers, quand vous êtes en communication avec une « personne d'intérêt ». Si le logiciel détecte qu'une personne est en train de tester le téléphone, de pratiquer une lecture de code ou un *reverse engineering* sur une partie de code pouvant révéler l'algorithme dormant, il peut substituer un processus de chiffrement en générant une nouvelle clé et en utilisant un générateur de nombres aléatoires malicieux pour maintenir par exemple, un accès furtif au terminal.

- Ceci nécessitera peut-être la coopération du fabricant de terminal, afin que le logiciel embarqué puisse bénéficier d'une seconde couche d'opacité intelligente : celle qui aura été mise en place par le fabricant de terminal pour en dissimuler l'existence (qui se traduit dans les faits par des « mécanismes d'ignorance » qui instruisent le terminal à *ne pas détecter* la présence d'un logiciel bénéficiant d'accès administrateurs incongrus).

Le parallèle entre l'affaire Volkswagen et un scénario d'interception et de déchiffrement de communications téléphoniques n'est pas une métaphore. Il est bien réel. L'utilisation d'un générateur de nombres aléatoires n'est pas un exemple fictionnel : la possibilité de son usage est déjà intégrée dans la publication en 2013 d'un standard cryptographique de la NIST[18]. En septembre

[18] Stephen Checkoway, Matthew Fredrikson, Ruben Niederhagen, Adam Everspaugh, Matthew Green, Tanja Lange, Thomas Ristenpart, Daniel J. Bernstein, Jake Maskiewicz, and Hovav Shacham, Daniel J. Bernstein, Jake Maskiewicz, and Hovav Shacham « On the Practical Exploitability of Dual EC in TLS Implementations », *Usenix Security Symposium*, 20-24 août 2014, San Diego, CA. http://dualec.org/DualECTLS.pdf.

2013, le *New York Times* alertait le public sur un programme visant à influencer les standards cryptographiques afin d'obtenir par *backdoor* un accès pré-cryptographique à l'échange de messages, en cherchant à généraliser un principe légal de « déchiffrement opportuniste », c'est-à-dire, à l'instar de l'exemple de Volkswagen, la capacité de déclencher à distance, par apprentissage algorithmique, le déchiffrement « opportun » d'une partie des messages d'une cible, afin d'éviter la détection[19]. Avec une approche différente, la Chine utilise la détection automatique de mots-clés sur l'intégralité des systèmes de messagerie instantanés sur son territoire, permettant l'interception des messages « opportuns » pour la RPC[20].

La vulnérabilité de sécurité que nous discutons ici ne requiert pas le déploiement d'une « algorithmique d'État ». Elle ne repose ni sur un niveau de compétence particulièrement élevée, ni sur une logistique de déploiement uniquement réservée aux États. Le petit exemple que nous vous présentons ci-dessous, sous la forme d'un algorithme d'une grande simplicité, peut entièrement remplir une telle fonction, et retirer toute garantie de confidentialité d'un échange, si vous échouiez à le détecter :

```
if (isInBlacklist (phoneNumber) || forensicsDetected ()) {
   const ephPrivKey = DUAL_EC(32)
}
else {
   const ephPrivKey = DevUrandom(32)
}
const ephPubKey = DH25519 (g, ephPrivKey)
sendEncryptedMessage(
   [msg, ephPrivKey, ephPubKey],
   phoneNumber)
```

L'argument opposé par les agences nationales, invoquant le risque de diffusion de codes pouvant aboutir dans les mains de l'adversaire (argument récurrent des agences d'État américaines) est donc fallacieux. Il n'y a rien dans de telles lignes de code qui ne puisse pas être réalisé par un hacker débutant, ou un jeune diplômé en informatique (le diplôme n'étant ici absolument pas

[19] Nicolle Perlroth, Jeff Larson et Scott Shanesept, « N.S.A. Able to Foil Basic Safeguards of Privacy on Web », *The New York Times*, 5 sept. 2013.

[20] Citizen LAb, « Asia Chats: Analyzing Information Controls and Privacy in Asian Messaging Applications », 4 nov. 2013. https://citizenlab.org/2013/11/asia-chats-analyzing-information-controls-privacy-asian-messaging-applications/.

nécessaire !). Cet algorithme est un simple « if.. then » qui instruit le logiciel de cryptographie de changer la clé de chiffrement si une personne est sur une liste noire. Si nous savons que le code du logiciel de messagerie ou du terminal téléphonique ne peut jamais être inspecté, alors nous pourrions installer ce petit sniplet de code dans votre application de messagerie chiffrée, et détourner vos échanges.

Si ces mêmes lignes de code étaient offusquées – c'est-à-dire noyées par un ajout de code contextuel masquant leur réel propos, dans un scénario d'inspection non autorisée, elles échapperaient probablement à leur détection pour un certain temps. L'offuscation de code n'est pas non plus une technique particulièrement sophistiquée. Elle est enseignée dans la plupart des écoles d'ingénieur.

Les deux cas présentés présentent plusieurs similarités : il n'y a pas de dangerosité dans la technique utilisée, ni d'éligibilité des techniques utilisées au statut de secret industriel ou d'intérêt national, rendant totalement injustifiable une opacité pour de tels motifs ; les deux approches reposent sur un algorithme distribué dont l'usager est en totale ignorance ; et les algorithmes embarqués ont une même fonction de pervertissement de l'apprentissage (par substitution de résultats, par amoindrissement ou élimination de la sécurité). Ce que Volkswagen a déployé pour tromper le régulateur, et le public, est déployé de la même façon et à échelle globale dans le chiffrement des communications privées.

L'ARGUMENT ECONOMIQUE EST-IL VALIDE ?

Ces deux cas posent donc le problème général de l'inspection de code au sein de sociétés fortement dépendantes de gouvernances algorithmiques distribuées. Les gouvernements aussi bien que les industriels opposent généralement l'argument que les *backdoors* (un accès distant à l'algorithme distribué) sont nécessaires pour des motifs soit de sécurité d'opération (maintenance et vérification à distance), soit des motifs légaux (leur responsabilité quant à la performance des algorithmes étant engagée), ou encore des motifs de sécurité nationale ou de propriété industrielle (la connaissance du code pouvant mener à sa perversion ou à son détournement). En réalité, dans les deux exemples que nous développons, c'est l'industriel lui-même, ou le gouvernement en coopération avec les industriels dans le cas du chiffrement des messages téléphoniques, qui se sont engagés dans une stratégie de perversion et d'assombrissement des codes, pour échapper à la détection des usagers, des équipementiers de tests et des usagers. Ce faisant, ils ont précisément contribué à une perturbation globale de l'ordre public, dont l'invocation (sécurité publique,

sécurité nationale, sécurité des véhicules) fut invoquée pour justifier l'opacité des dispositifs algorithmiques déployés.

L'argument de dernier recours opposé à une vérification transparente des codes des logiciels est économique : le coût d'une vérification de la fiabilité des logiciels serait trop élevé, si elle devait être effectuée non plus en amont mais *à l'aval* des chaînes de valeur (dans nos deux exemples : par tout tiers sur un terminal pendant son usage, par une autorité publique ou par transparence des codes sur des bancs de test de contrôle de pollution). De plus, un régime de transparence forcé sur l'industrie du logiciel serait, selon cet argument, destructeur de rentes, et menacerait à long terme l'industrie du logiciel soutenant l'économie numérique. Nous pourrions sans doute ajouter, comme nous l'évoquions au début de cet article, qu'une telle vérification ouverte, libre, externe et permanente et distribuée anéantirait toute stratégie d'asymétrie de connaissance, sur laquelle reposent justement les rentes des nouveaux grands conglomérats de l'industrie du numérique. Car au-delà de la vérification mathématique du code, les deux exemples présentés posent la question du « propos algorithmique », c'est-à-dire l'assurance que l'algorithme poursuit le but qui est annoncé, dans un ensemble d'opérations finies exclusivement dédiées à ce but, et ne produisant aucun résultat ayant des objectifs étrangers à ce but à l'insu de l'usager, de l'espace public, ou dans le cas des communications chiffrées, du territoire national dans lequel est déployé l'algorithme (ex. : souveraineté).

L'argument économique invoqué est ainsi double : d'une part, on retrouve l'argument classique opposant une économie de biens communs et une économie de marché (l'économie de marché faisant l'objet d'un postulat concernant sa capacité supérieure à soutenir des régimes d'innovation) ; et d'autre part, l'invocation – fallacieuse –, qu'une méthode de chiffrement dans le temps est dépendante de coûts de production non adressables localement, ignorant la simplicité du défi mathématique posé, et l'existence de solutions dans l'état de l'art. L'obstacle économique à la mise en œuvre d'un cadre de régulation autorisant une transparence locale résiderait donc sur les coûts associés à la vérification, qui ne pourrait être organisée de manière efficiente en dehors d'une organisation industrielle (pouvant lui appliquer des économies d'échelle, et étant garantie par des processus de certification), ou d'une organisation d'État (qui bénéficie de la concentration de moyens et de son impartialité *a priori*).

Les capacités algorithmiques contemporaines ne résident pas uniquement dans la qualité mathématique et analytique de leurs prémisses ou de leurs primitifs, mais dans la capacité à résister au cassage de leur protection. Le

problème de la cryptographie ne réside pas dans la limite de fonctions calculables, dont on sait l'infinité, mais bien dans le coût exponentiel d'élévation des protections, et de la surenchère algorithmique, aussi bien en taille de données traitées (Big Data), qu'en complexité algorithmique (pour défaire et se prémunir des attaques). Le chiffrement asymétrique, par exemple, repose sur un partage de clés publiques (à la disponibilité de tous), et l'usage d'une clé privée de déchiffrement qui reste secrète. La vérification légitime du chiffrement n'est pas consommatrice d'une capacité de calcul (CPU) qui ne soit pas réalisable localement, même sur des objets connectés. C'est l'opération inverse, celle qui consiste à briser un chiffrement, qui est fortement consommatrice de ressources.

Le chiffrement asymétrique est construit sur des notions mathématiques librement disponibles et dont l'implémentation est facile : Si Alice et Bob disposent de clefs privées *a* et *b*, un générateur commun *g* et un nombre premier commun *p*, le calcul de leurs clefs publiques sera les simples exponentiels (g^a, g^b) et leur secret commun de chiffrement est :

$$(g^b)^a = (g^a)^b$$

Toutes les exponentiels se calculent modulo p.

Cette simplicité mathématique, établie in 1976 par Diffie et Hellman[21], a rendu la cryptographie moderne difficile à contrôler ; la preuve résidant dans l'échec de la tentative, par les États-Unis, de contrôler l'exportation de logiciels cryptographiques comme « munitions » en 1995. La tradition de recherche en cryptographie assure que toute nouvelle construction obtient sa sécurité en se basant sur un problème mathématique non résolu, et non sur la confidentialité de ses méthodes : au contraire, la méthodologie est publiée en détail pour obtenir l'examen rigoureux des autres cryptographes.

La sécurité cryptographique, aussi bien que sa loyauté au sens mathématique (sa possibilité de vérification), reposent, pour l'instant, sur le problème de la factorisation d'entiers, et sur le problème de logarithme discret. Les meilleurs algorithmes de factorisation peuvent pousser les limites de la factorisation jusqu'à un maximum théorique de 1 024 octets, mais nécessitent alors plusieurs mois de calculs avec une consommation de puissance ne pouvant assurer leur usage quotidien, voire même une application industrielle raisonnable. Donc, pour qu'un « surveillant », en utilisant une terminologie foucaldienne, puisse briser la

[21] Diffie, W., Hellman, M. (1976). "New directions in cryptography" (PDF). IEEE Transactions on Information Theory 22 (6): 644–654. https://ee.stanford.edu/%7Ehellman/publications/24.pdf.

confidentialité d'un chiffrement, il doit premièrement résoudre une instance de problème de factorisation ou de logarithme discret.

Les progrès réalisés en termes de limite de factorisation sont dus à la fois aux avancées récentes en termes de formalisation, à la montée en puissance et vitesse des processeurs, des mémoires, et des bus d'accès au stockage, mais aussi à l'innovation des algorithmes qui ne nécessitent pas des champs finis aussi larges, comme la cryptographie avec les courbes elliptiques. Le chiffrement et le déchiffrement par les communicants sont donc économiquement faciles, mais existe-t-il pour le *surveillant* une limite pratique, à dimension humaine, de la capacité de factorisation « soutenable » d'un point de vue économique ?

Si nous prenons l'hypothèse d'une généralisation globale du chiffrement avec un taux de pénétration de 70 % à l'échéance 2030, en supposant un gain relatif de population (incertain à ce jour), une question se pose : quelle capacité de calcul globale (« terrestre ») faudrait-il produire pour qu'un surveillant puisse pénétrer des interactions quotidiennes de 6 milliards d'individus ? La réponse : il est infiniment plus simple mathématiquement de chiffrer que de déchiffrer par la force. Un chiffrement qui prend quelques nanosecondes pour un utilisateur légitime rend impossible, en théorie, sa résolution par force (pour le briser) si le problème y étant associé est mathématiquement rigoureux.

C'est exactement pour cette raison qu'on observe des subversions des garanties de sécurité dans de nombreux secteurs privés, de Volkswagen à Google, qui dépendent de la désactivation subreptice des garanties de sécurité au niveau du logiciel *avant* qu'il ne puisse rendre les données confidentielles. Ces stratégies, observées de façon sporadique et imprévisible dans une grande part du secteur privé, donnent naissance à l'argument que la sécurité des intérêts, économiques ou sociétaux, dépend de l'implémentation de techniques de *vérification formelle publiquement disponibles* de systèmes de sécurité de toutes sortes, et non sur des tests « *black box* » qui se produisent sans qu'un examen rigoureux et vérifiable par un tiers ne soit possible. Au-delà même de la vérification directe des codes sources, et de la loyauté de leurs propos et de leur programmation, la science algorithmique elle-même a réalisé suffisamment de progrès pour pouvoir estimer, par des lois de Poisson, de manière probabiliste le délai avant lequel un ensemble de codes logiciels présentera un défaut de fiabilité[22]. Il n'y a aucune raison pour qu'une approche similaire ne puisse être développée, pour tester l'existence d'un

[22] Cf. Néstor R. Barraza, « Parameter Estimation for the Compound Poisson Software Reliability Model », *International Journal of Software Engineering and its Applications. IJSEIA* Vol. 7, No.1, 201, et avec une différente approche : N. R. Barraza (2015) « A Parametric Empirical Bayes Model to Predict Software Reliability Growth », *Procedia Computer Science*, Vol. 62 p. 360-369.

backdoor, d'un détournement de propos ou de finalité, ou le comportement incongru d'un composant logiciel, sans interférer avec la totalité du secret logiciel ou industriel. La possibilité d'un algorithme de vérification, transparent dans sa nature et sa construction, et son accès libre à l'algorithme distribué (contrôle de pollution, chiffrement de communications, etc.) est bien le seul enjeu de régulation dont l'impact positif sur l'ordre public, la défense des libertés fondamentales, peut être lui-même vérifié et régulé avec la sérénité qu'aucune subversion de propos algorithmique ait été mise en œuvre.

La base d'une telle procédure de vérification scientifiquement légitime se base sur des notions formelles, définies dans la science de la cryptographie :

- Le principe de Kerckhoffs[23], qui assume toujours que l'adversaire connaît tous les détails du système de chiffrement ou de sécurité, sauf la clef. Ce principe est fondamental à toute procédure scientifique de cryptographie depuis le XIX^e^ siècle. Le respect de ce principe est ce qui permet à la communauté cryptographique et mathématique de vérifier le travail scientifique dans le domaine.

- Tout protocole de sécurité cryptographique doit se baser sur des primitifs de chiffrement standardisé par des corps scientifiques, et dont la solidité mathématique est indépendamment vérifiable. Ceci évite l'utilisation dangereuse et définitivement non nécessaire de primitifs uniques à une industrie, voire sans vérification externe.

- Les protocoles cryptographiques doivent être formellement spécifiés, ceci incluant les buts, les hypothèses et le modèle de menaces contre quoi il est censé de protéger. Le code doit être fonctionnellement adhérant à la spécification.

- La capacité d'obtenir à tout moment (durant et après la certification, et même après l'entrée dans le marché), une vérification de l'intégrité du système. Dans le cas où la confidentialité est strictement nécessaire, le système de sécurité peut offrir un système de preuve de classe « *zero knowledge* », ce qui permet d'obtenir une vérification par preuve d'intégrité, tout en permettant au système de garder des informations secrètes.

CONCLUSION

Au lieu de dépendre des tests que Volkswagen conduit dans ses laboratoires, les garanties de sécurité et d'émissions auraient dû être déjà implémentés avec une logique qui permet une vérification formelle externe immédiate. Cette

[23] Auguste Kerckhoffs, « La cryptographie militaire », *Journal des sciences militaires,* vol. IX, p. 5–38, jan. 1883, pp. 161–191, févr. 1883 https://www.bibnum.education.fr/sites/default/files/kerckhoffs-texte.pdf.

méthodologie fait déjà partie de l'état de l'art usuel dans le domaine de la cryptographie. Le secteur privé, aussi bien que les secteurs gouvernementaux, pourraient bénéficier de cette rigueur déjà solidement définie et mathématiquement comprise.

Notre analyse suggère que la relation entre « algorithme » et « ordre public » ne peut être réduite à une simple question de certification et de vérification normative exercée par le régulateur ou par l'industriel. Dans les deux cas présentés, les algorithmes incorporés aux équipements ont été pervertis par leurs concepteurs (Volkswagen) ou par l'écosystème (autorité régulatrice – industriels), ayant pourtant passé les tests de certification et de vérification en amont de leur implémentation. Dans les deux cas, la capacité algorithmique a été détournée afin d'échapper à la détection en utilisant des systèmes d'apprentissage situationnel délibérément instruits pour éviter ces contrôles (par la substitution de clés pour le chiffrement des communications, par la détection de l'environnement de tests et le lancement d'une routine de test factice dans l'automobile).

La décision récente de la Bibliothèque du Congrès d'intégrer de nouvelles exemptions dans le *Digital Millenium Copyright Act* (DMCA)[24], de donner accès aux utilisateurs d'un véhicule au logiciel qui régit la conduite et le comportement de leurs véhicules a été soutenue par l'administration nationale des télécommunications et de l'information (NTIA) sous le motif qu'une telle exemption « était nécessaire afin de permettre aux consommateurs de perpétuer une longue pratique de pouvoir travailler sur leurs propres véhicules »[25]. Ces exemptions, dans la note finale de la DMCA, concernent également la pratique qui consiste à déverrouiller le logiciel d'un téléphone portable (« *jailbreak* » ; *Unlocking Act*)[26], et s'étend, dans ces exemptions, à l'ensemble des objets connectés. L'argument invoqué par l'Electronic Frontier Foundation (EFF) pour la révision du DMCA est celui de la restriction de choix (limitation de la concurrence), la détérioration des actifs économiques (le verrouillage accélérant

[24] Library of Congress, U.S. Copyright Office, 37 CFR Part 201, [Docket No. 2014-07], Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies, Final Rule, http://copyright.gov/1201/2015/fedreg-publicinspectionFR.pdf.

[25] Sean O'Kane, « Automakers just lost the battle to stop you from hacking your car. The vice grip on car software gets forcibly loosened », *The Verge*, 27 oct. 2015.

[26] *« Proposed Class 12: This proposed class would allow the unlocking of all-purpose tablet computers. This class would encompass devices such as the Apple iPad, Microsoft Surface, Amazon Kindle Fire, and Samsung Galaxy Tab, but would exclude specialized devices such as dedicated e-book readers and dedicated handheld gaming devices. », op. cit.* p. 29.

l'obsolescence), et la protection de l'environnement (l'obsolescence artificielle engendrant une pollution bien réelle). Mais le DMCA n'a retenu ces exemptions que pour les terminaux usagés, excluant les terminaux neufs, et excluant aussi les logiciels embarqués dans l'électronique des véhicules motorisés (*op. cit.*, p. 32).

La Classe 16 du nouveau DMCA, cependant, autorise la désinstallation « de logiciels pré-installés non désirés d'un terminal »[27], ce qui permet légalement aux États-Unis, de déverrouiller un terminal communiquant portable neuf ; et d'en retirer des composants logiciels potentiellement pervertis. Le lobby industriel de l'automobile a cependant obtenu que l'électronique communicante embarquée des véhicules à moteur soit exclue de cette disposition. Et, sans surprise, les mêmes mécanismes de défense ont été déployés avec succès par les lobbies des terminaux télévisuels, des équipementiers de la santé, à l'exception du *reverse engineering* pour des motifs de sécurité, qui lui bénéficie de la bienveillance du régulateur ! [28]. Le fondement de cette dernière exemption concerne la recherche de sécurité exercée dans une « démarche de bonne foi ». Elle vise donc à autoriser l'ingénierie inverse d'un code ou d'un algorithme, et donc l'éventuel cassage de son chiffrement, dans l'objectif « d'une recherche de sécurité de bonne foi, c'est-à-dire accéder à un programme d'ordinateur pour le seul but d'un test de bonne foi, d'investigation et/ou de correction d'une faille de sécurité ou une vulnérabilité, où cette activité peut être conduite dans un environnement contrôlé, conçu pour éviter tout dommage à un individu ou au public, et où l'information n'est recueillie que pour l'activité primaire de la promotion de la sécurité et de la sûreté de cette classe de machines, et non dans une manière qui pourrait encourager la violation de copyright »[29].

Questionner le « propos algorithmique » et la loyauté d'un algorithme est donc devenu légalement possible le 27 octobre 2015 aux États-Unis**.** Interdire l'inspection d'un code, et s'autoriser son obscurcissement, est aussi légalement possible sur des motifs de violation potentielle du copyright. Mais le nouveau texte de la DMCA laisse encore supposer que l'existence d'une propriété intellectuelle est en opposition avec une possibilité d'inspection des codes

---

[27] *Proposed Class 16: This proposed class would permit the jailbreaking of wireless telephone handsets to allow the devices to run lawfully acquired software that is otherwise prevented from running, or to remove unwanted preinstalled software from the device*. (p. 33).

[28] *« Computer programs, where the circumvention is undertaken on a lawfully acquired device or machine on which the computer program operates solely for the purpose of good-faith security research and does not violate any applicable law, including without limitation the Computer Fraud and Abuse Act of 1986, as amended and codified in title 18, United States Code; and provided, however, that, except as to voting machines, such circumvention is initiated no earlier than 12 months after the effective date of this régulation* [...] », p. 50.

[29] *Op. cit.*, p. 51.

sources. Ceci n'est plus une réalité dans les faits. La société Apple, depuis le 29 octobre 2015, a décidé de communiquer en source libre ses standards cryptographiques, incluant leurs primitifs cryptographiques, au public, afin que tout un chacun puisse inspecter librement le code[30]. Il s'agit, comme l'indique Apple, des mêmes bibliothèques de codes cryptographiques que la société a soumises pour son examen de conformité auprès du FIPS 140-2 Level 1 (US Federal Information Processing Standards). Pour autant, Apple n'a pas révélé ses secrets industriels, n'a pas compromis sa propriété intellectuelle, et n'a pas souffert d'un revers de marché ou d'une perte de performance économique.

Au-delà des implications morales et philosophiques qu'une économie algorithmique peut soulever, et auxquelles nul ne peut faire objection, la relation entre l'algorithme et l'ordre public donne le sentiment que le régulateur est faible et désœuvré ; aussi bien vis-à-vis des industriels des MEAs, que de l'État lorsque celui-ci introduit, comme on vient de le voir pour les États-Unis, une disposition légale lui permettant d'éliminer de façon discrète un chiffrement au mépris du respect de la vie privée et du droit international.

Que ce soit au motif de l'ordre public, à celui de la défense des libertés humaines fondamentales, ou celui du simple droit de tout un chacun de savoir de quelles « variables », de quelles « règles », et avec « quelles données » est géré son véhicule, sa messagerie, l'interaction avec ses amis, l'inspection du propos et de la fonction d'un algorithme doit être *toujours possible*. Nous avons montré dans cet article que les mathématiques étaient capables d'estimer la fiabilité d'un logiciel, la cohérence de son comportement vis-à-vis de ses buts affichés, sans forcément en compromettre ni la propriété intellectuelle, ni l'ordre public économique qu'il soutient. Il revient au régulateur, pour le bien être public, et pour éviter des opérations de tromperie comme celle de Volkswagen, ou celle de la norme de chiffrement nord-américaine, de cesser d'être l'invité impromptu de cette conversation, et d'en devenir l'hôte.

[30] https://developer.apple.com/cryptography/.