Communication numérique

Algorithmes – Régulation – Droits fondamentaux

La CJUE dessine le noyau dur d'une future régulation des algorithmes

CJUE 6 octobre 2020, aff. C-511/18
La Quadrature du Net

L'article 15, paragraphe 1, de la directive 2002/58/CE (directive vie privée et communications électroniques), telle que modifiée par la directive 2009/136/CE, lu à la lumière des articles 7, 8 et 11 ainsi que de l'article 52, paragraphe 1, de la Charte des droits fondamentaux, doit être interprété en ce sens qu'il ne s'oppose pas à une réglementation nationale imposant aux fournisseurs de services de communications électroniques de recourir, d'une part, à l'analyse automatisée ainsi qu'au recueil en temps réel, notamment, des données relatives au trafic et des données de localisation et, d'autre part, au recueil en temps réel des données techniques relatives à la localisation des équipements terminaux utilisés, lorsque :

- le recours à l'analyse automatisée est limité à des situations dans lesquelles un État membre se trouve confronté à une menace grave pour la sécurité nationale qui s'avère réelle et actuelle ou prévisible, le recours à cette analyse pouvant faire l'objet d'un contrôle effectif, soit par une juridiction, soit par une entité administrative indépendante, dont la décision est dotée d'un effet contraignant, visant à vérifier l'existence d'une situation justifiant ladite mesure ainsi que le respect des conditions et des garanties devant être prévues, et que
- le recours à un recueil en temps réel des données relatives au trafic et des données de localisation est limité aux personnes à l'égard desquelles il existe une raison valable de soupçonner qu'elles sont impliquées d'une manière ou d'une autre dans des activités de terrorisme et est soumis à un contrôle préalable, effectué, soit par une juridiction, soit par une entité administrative indépendante, dont la décision est dotée d'un effet contraignant, afin de s'assurer qu'un tel recueil en temps réel n'est autorisé que dans la limite de ce qui est strictement nécessaire. En cas d'urgence dûment justifiée, le contrôle doit intervenir dans de brefs délais.

Commentaire

Winston Maxwell
Director of Law
& Technology Studies
Télécom Paris - Institut
Polytechnique de Paris

La loi n° 2015-912 du 24 juillet 2015 relative au renseignement permet aux autorités de renseignement d'imposer aux opérateurs de communications électroniques et aux hébergeurs la mise en place d'algorithmes pour détecter d'éventuelles activités terroristes [1]. L'algorithme, dont les paramètres sont spécifiés par l'autorisation du Premier ministre après avis de la Commission national de contrôle des techniques de renseignement (CNCTR), analyse en temps réel des données de connexion et de géolocalisation, mais pas le contenu des communications [2].

La Quadrature du Net a contesté la légalité du décret d'application [3] de cette loi devant le Conseil d'État, et celui-ci a posé une question à la CJUE sur la compatibilité de ce dispositif avec la directive 2002/58/CE et la Charte des droits fondamentaux de l'Union européenne (la Charte) [4]. Même si cette affaire concerne la lutte contre le terrorisme et les données de connexion, l'arrêt de la Cour [5] nous permet de dégager quelques principes qui s'appliqueront de manière plus large [6] à tout algorithme ayant des impacts défavorables sur les droits et libertés individuels, et notamment les algorithmes utilisés pour lutter contre diverses formes de criminalité et de contenus illicites : terrorisme, blanchiment d'argent, cyberattaques, fraude fiscale, fraude à la sécurité sociale, diffusion d'images pédopornographiques, désinformation, contenus haineux, contrefaçon de droit d'auteur. Que ce soit un algorithme de filtrage pour bloquer des téléchargements d'œuvres protégées

1 Art. 5 de la loi n° 2015-912 du 24 juill. 2015 a créé l'art. L. 851-3 CSI.
2 Pour une description complète du dispositif, v. le rapport annuel 2018 de la CNCTR, p. 16 à 19, et le rapport des députés Didier Paris et Loïc Kervan, enregistré à l'Assemblée nationale le 8 juill. 2020, p. 16 à 25.
3 Décr. n° 2016-67 du 27 janv. 2016 relatif aux techniques de recueil de renseignement.
4 CE 26 juill. 2018, n° 394922, *Quadrature du Net*, Lebon ; AJDA 2018. 1586 ; *ibid.* 2027, note F.-X. Bréchot ; D. 2018. 1756, obs. M.-C. de Montecler ; RTD eur. 2019. 541, obs. A. Bouveresse. Les recours de la Quadrature du Net concernaient plusieurs décrets, dont ceux précisant les conditions de conservation des données par les opérateurs de communications électroniques et les hébergeurs. Nous évoquerons seulement l'aspect du litige concernant les algorithmes destinés à détecter des activités terroristes.
5 CJUE 6 oct. 2020, aff. jtes C-511/18, C-512/18 et C-520/18, *La Quadrature du net (Assoc.)*, AJDA 2020. 1880 ; D. 2020. 1948, et les obs. ; *ibid.* 2262, obs. J. Larrieu, C. Le Stanc et P. Tréfigny ; AJ pénal 2020. 531. Notre analyse se limitera à l'affaire C-511/18.
6 Sur le champ d'application de l'arrêt, voir notre analyse à la section III.

par le droit d'auteur [7], ou pour alerter les autorités sur un risque de terrorisme, la grille d'analyse en droits fondamentaux sera similaire.

Notre analyse commencera par un rappel des différentes formes d'algorithmes utilisés dans la détection et la prédiction d'activités illicites (I). Nous analyserons ensuite le raisonnement de la Cour et les sept principes de régulation qui en découlent (II). Enfin, nous examinerons le champ d'application de l'arrêt pour conclure qu'il s'appuie sur les principes de la Charte et aura vocation à s'appliquer à tout algorithme ayant des effets indésirables sur les droits et libertés individuels (III).

I - LES DIFFÉRENTS TYPES D'ALGORITHMES

Les algorithmes se divisent en deux grandes familles : la famille de l'IA symbolique et la famille de l'apprentissage statistique, aussi appelée *machine learning*. L'IA symbolique s'appuie sur des règles de logique et des connaissances humaines. Les algorithmes issus de l'apprentissage statistique (*machine learning*) s'appuient sur des règles créées par l'algorithme d'apprentissage lui-même, après l'analyse d'une grande masse de données. À l'intérieur de la famille *machine learning* on distingue l'apprentissage supervisé et l'apprentissage non-supervisé. L'apprentissage supervisé consiste à entraîner l'algorithme sur des données étiquetées [8]. L'apprentissage non-supervisé consiste à laisser un algorithme analyser une grande masse de données non-étiquetées pour essayer de créer des groupes (*clusters*) de données ayant des points communs. L'apprentissage supervisé fonctionne bien lorsqu'il existe beaucoup d'exemples étiquetés pour entraîner l'algorithme. En revanche, l'apprentissage supervisé aura du mal à construire un modèle statistique pour prévoir des événements rares, car il n'existera pas assez d'exemples étiquetés pour entraîner le modèle. La criminalité et le terrorisme étant des événements dont la détection est plutôt rare, l'apprentissage non-supervisé sera plus adaptée car cette technique permet de créer des groupes homogènes d'événements (opérations bancaires, communications) et identifier des événements qui sortent du lot (*outliers*), et qui peuvent correspondre à une activité criminelle. Les trois approches – IA symbolique, apprentissage supervisé, apprentissage non-supervisé – peuvent être combinées dans des approches dites hybrides.

Quelle que soit l'approche, les algorithmes de prédiction engendront des erreurs, à la fois des faux positifs (un événement identifié à tort comme suspicieux), ou des faux négatifs (un événement criminel non-détecté). Même un algorithme destiné à bloquer le téléchargement d'œuvres protégées par le droit d'auteur aura un certain nombre de faux positifs car il sera incapable d'identifier des exceptions au droit d'auteur (parodie, critique, etc.) [9]. Les systèmes de détection de contenus haineux ou de désinformation souffriront du même défaut, conduisant à un certain niveau de sur-blocage. Les algorithmes de *machine learning* sont probabiliste par nature. Leurs résultats se traduisent en niveau de probabilité statistique, par exemple 88 % de probabilité qu'il s'agit d'un contenu illégal. Les systèmes de règles logiques (IA symbolique) sont déterministes, ils sortent des résultats certains – oui ou non. Malgré leur apparence de certitude, les systèmes déterministes conduisent à des faux positifs, car la question posée à l'algorithme ne correspondra pas généralement à la vraie question qui nous intéresse. Au lieu de poser la question : « Est-ce que ce client participe à un réseau de blanchiment ? », on va poser une question plus simple : « est-ce que ce client a effectué un dépôt d'argent en liquide supérieur à 5 000 € ? » La deuxième question est simple, mais ne répondra qu'imparfaitement à la première question, d'où l'existence de faux positifs, même au sein d'un algorithme déterministe.

Même un algorithme destiné à bloquer le téléchargement d'oeuvres protégées par le droit d'auteur aura un certain nombre de faux positifs car il sera incapable d'identifier des exceptions au droit d'auteur (parodie, critique, etc.)

Le fonctionnement d'un algorithme fondé sur des règles logiques (IA symbolique) sera généralement plus compréhensible que le fonctionnement d'un algorithme de *machine learning* car les règles sont définies au départ par un humain. Cependant, de nouveaux outils tentent de combler cette lacune pour rendre les algorithmes de *machine learning* plus interprétables [10].

II - L'ANALYSE DE LA CJUE DES ALGORITHMES DE DÉTECTION D'ACTIVITÉS TERRORISTES

L'algorithme prévu par l'article L. 851-3 du code de la sécurité intérieure analyse en temps réel de grandes masses de données de trafic et de localisation pour détecter des « individus dont les comportements, notamment compte tenu de leurs modes de communication, sont susceptibles de révéler une menace terroriste » [11]. La question de la compatibilité de ces algorithmes avec la directive 2002/58/CE et la Charte est posée parce que le dispositif effectue un *traitement général et indifférencié* de l'ensemble des données de tous les utilisateurs des

7 CJUE 24 nov. 2011, aff. C-70/10, *Sté Scarlet Extended c/ Société belge des auteurs, compositeurs et éditeur SCRL*, D. 2011. 2925, obs. C. Manara ; *ibid*. 2012. 2343, obs. J. Larrieu, C. Le Stanc et P. Tréfigny ; *ibid*. 2836, obs. P. Sirinelli ; Légipresse 2011. 657 et les obs. ; *ibid*. 2012. 167, comm. O. Bustin ; RSC 2012. 163, obs. J. Francillon ; RTD eur. 2012. 404, obs. F. Benoît-Rohmer ; *ibid*. 957, obs. E. Treppoz.

8 Par ex. une série de pixels sera préalablement étiquetée comme correspondant à l'image d'un chat.

9 CJUE 24 nov. 2011, aff. C-70/10, *Sté Scarlet Extended c/ Société belge des auteurs, compositeurs et éditeur SCRL*, préc., pt 52.

10 V. Beaudouin et al., Identifying the 'Right'Level of Explanation in a Given Situation (May 13, 2020). Proceedings of the First International Workshop on New Foundations for Human-Centered AI (NeHuAI), Santiago de Compostella, Spain, September 4, 2020, CEUR Workshop Proceedings, vol. 2659, p. 63.

11 CJUE 6 oct. 2020, aff. jtes C-511/18, C-512/18 et C-520/18, *La Quadrature du net (Assoc.)*, préc., pt 65.

réseaux de communications électroniques, une approche *a priori* excessive et contraire à la Charte compte tenu de la jurisprudence récente de la Cour [12].

Dans son analyse, la Cour commence par caractériser le niveau d'ingérence avec les droits fondamentaux. Selon la CJUE, l'ingérence est « particulièrement grave » [13], puisque l'algorithme analyse l'ensemble des données de trafic et de localisation de toute la population. Les droits affectés sont non seulement la protection de la vie privée et des données à caractère personnel mais aussi la liberté d'expression, car les techniques de surveillance intrusives ont un effet dissuasif sur l'expression [14]. Compte tenu de ce niveau d'ingérence particulièrement élevé, le déploiement de l'algorithme ne peut être envisagé que de manière exceptionnelle pour faire face à une « menace grave pour la sécurité nationale qui s'avère réelle et actuelle ou prévisible » [15]. De plus, l'autorisation de mise en œuvre de cette technique de surveillance doit « faire l'objet d'un contrôle effectif soit par une juridiction, soit par une autorité administrative indépendante, dont la décision est contraignante, visant à vérifier l'existence d'une situation justifiant ladite mesure ainsi que le respect des conditions et garanties devant être prévues » [16].

Les droits affectés sont non seulement la protection de la vie privée et des données à caractère personnel mais aussi la liberté d'expression, car les techniques de surveillance intrusives ont un effet dissuasif sur l'expression.

La Cour spécifie ensuite les conditions auxquelles doivent répondre les algorithmes. Ces conditions apparaissaient déjà dans l'avis de 2017 sur l'accord UE-Canada sur les données de passagers aériens (PNR) [17] : « les modèles et critères préétablis sur lesquels se fonde ce type de traitement de données doivent être, d'une part, *spécifiques et fiables*, permettant *d'aboutir à des résultats* identifiant des individus à l'égard desquels pourrait peser un soupçon raisonnable de participation à des infractions terroristes, et d'autre part, *non-discriminatoires* » [18]. Nous examinerons chacun de ces critères successivement.

Une première interrogation concerne les mots « *modèles et critères préétablis* » [19]. Une fois mis en production, le modèle, et le choix des données d'entrée sur lequel il repose (les critères), seront « *préétablis* », quel que soit le type d'algorithme. Même un modèle issu du *machine learning* ne change pas une fois terminée la phase d'apprentissage. On peut s'interroger, cependant, sur un modèle de machine *learning* qui est ré-entraîné périodiquement. Est-ce que le modèle restera « préétabli » dans ce cas ? Les modèles et critères doivent aussi être « *spécifiques* ». Aux yeux de la Cour, la spécificité vise probablement la transparence et l'interprétabilité du modèle, conditions nécessaires pour son contrôle effectif du modèle par un tribunal [20]. Un algorithme défini en termes généraux et dont les détails seraient non-spécifiés et évolutifs serait difficile à contrôler [21]. Mis ensemble, les termes « préétablis » et « spécifiques » signifient que les détails de l'algorithme doivent être décrits de manière détaillée et compréhensible dans l'autorisation, et que ces détails ne doivent pas changer au fil du temps.

Ensuite, le modèle doit être « *fiable* », à savoir qu'il doit conduire avec un certain niveau de performance à l'identification d'individus à l'égard desquels pourrait peser un soupçon raisonnable de participation à des infractions terroristes. L'efficacité à 100 % n'existe pas [22]. Le modèle générera un certain nombre de faux positifs, et laissera passer un certain nombre de faux négatifs, à savoir de vrais complots terroristes non-détectés par l'algorithme. Dans un dispositif de lutte contre le terrorisme, quel est le niveau acceptable de personnes soupçonnées à tort ? Selon le gouvernement américain, les algorithmes de détection de risque terroriste utilisés pour analyser les données de passagers aériens génèrent plus de 98 % de faux positifs [23]. Le problème est complexe car baisser le taux de faux positifs entraîne automatiquement une baisse du nombre de vrais positifs détectés. Le taux acceptable de faux positifs par rapport aux vrais positifs dépend du contexte, et notamment du niveau de préjudice subi par l'individu et par la société en raison de chaque type d'erreur. Dans la lutte antiterroriste, un faux positif signifie qu'une personne innocente sera soumise, peut-être à son insu, à une vérification plus approfondie, voire dans certains cas à une arrestation à tort [24]. Un faux négatif signifie qu'un vrai complot terroriste échappera au système, avec des conséquences graves pour la société. La Cour ne fournit pas d'indication sur le niveau tolérable de faux positifs. On

12 CJUE 8 avr. 2014, aff. jtes C-293/12 et C-594/12, *Digital Rights Ireland Ltd*, AJDA 2014. 773 ; *ibid*. 1147, chron. M. Aubert, E. Broussy et H. Cassagnabère ; D. 2014. 1355, et les obs., note C. Castets-Renard ; *ibid*. 2317, obs. J. Larrieu, C. Le Stanc et P. Tréfigny ; Légipresse 2014. 265 et les obs. ; RTD eur. 2014. 283, édito. J.-P. Jacqué ; *ibid*. 283, édito. J.-P. Jacqué ; *ibid*. 2015. 117, étude S. Peyrou ; *ibid*. 168, obs. F. Benoît-Rohmer ; *ibid*. 786, obs. M. Benlolo-Carabot ; CJUE 21 déc. 2016, aff. jtes C-203/15 et C-698/15, *Tele2 Sverige AB c/ Post-och telestyrelsen*, AJDA 2016. 2466 ; *ibid*. 2017. 1106, chron. E. Broussy, H. Cassagnabère, C. Gänser et P. Bonneville ; D. 2017. 8 ; *ibid*. 2018. 1033, obs. B. Fauvarque-Cosson et W. Maxwell ; Dalloz IP/IT 2017. 230, obs. D. Forest ; JAC 2017, n° 43, p. 13, obs. E. Scaramozzino ; RTD eur. 2017. 884, obs. M. Benlolo-Carabot ; *ibid*. 2018. 461, obs. F. Benoît-Rohmer ; Rev. UE 2017. 178, étude F.-X. Bréchot.

13 CJUE 6 oct. 2020, aff. jtes C-511/18, C-512/18 et C-520/18, *La Quadrature du net (Assoc.)*, préc., pt 174.

14 *Ibid.*, pt 118.

15 *Ibid.*, pt 177.

16 *Ibid.*, pt 179.

17 Avis 1-/15 Accord PNR UE-Canada du 26 juill. 2017.

18 *Ibid.*, pt 180.

19 L'adjectif « préétablis » s'applique à la fois aux modèles et aux critères. V. la version anglaise de l'arrêt qui utilise les mots : « *pre-established models and criteria* ».

20 Sur la nécessité d'explicabilité algorithmique pour permettre un contrôle effectif au regard de la CEDH, v. Tribunal du district de la Haye, 5 févr. 2020, aff. n° C-09-550982-HA ZA 18-388, *NJCM c/ le Gouvernement des Pays Bas*.

21 Cons. const., 12 juin 2018, n° 2018-765 DC, § 71, AJDA 2018. 1191 ; D. 2019. 1248, obs. E. Debaets et N. Jacquinot ; RTD eur. 2018. 830, obs. D. Ritleng.

22 CJUE 6 oct. 2020, aff. jtes C-511/18, C-512/18 et C-520/18, *La Quadrature du net (Assoc.)*, préc., pt 182.

23 Conseil de l'Europe, rapport de Douwe Korff et Marie Georges, Passenger Name Records, data mining & data protection : the need for strong safeguards, 15 juin 2015, T-PD(2015)11, p. 19.

24 *Ibid.*, p. 20.

sait seulement que la performance du système doit « permettre d'aboutir à des résultats identifiant des individus à l'égard desquels pourrait peser un soupçon raisonnable »[25]. Mais au bout de combien de tentatives erronées ? La Cour ne nous le dit pas. Dans un autre contexte – la lutte contre des contenus haineux par exemple – le taux acceptable de faux positifs serait sans doute différent.

La condition suivante imposée par la Cour concerne le caractère *non-discriminatoire* du dispositif : les modèles et critères doivent être non-discriminatoires[26] et,c ne doivent pas s'appuyer uniquement sur des données sensibles, telles que la couleur de peau ou l'origine ethnique. La Cour n'interdit pas tout traitement de ces données sensibles dans les modèles, ce qui peut surprendre[27], mais elle interdit tout modèle qui s'appuierait sur le postulat qu'une donnée telle que la religion ou l'origine ethnique pourrait par elle-même et indépendamment du comportement individuel de la personne être pertinente au regard de la prévention du terrorisme[28].

Construire un mécanisme de tests anti-discrimination peut se heurter aux principes du RGPD sur la non-utilisation des données sensibles.

Pour garantir le caractère non-discriminatoire du modèle, il faut un système de tests. Qu'ils soient construits à partir de règles logiques définies par des humains, ou par un mécanisme d'apprentissage statistique, les modèles produisent généralement des résultats biaisés. Ces biais ont de multiples causes[29], et pour les identifier, il faut prévoir des mécanismes de tests. Ces tests consisteraient à mesurer le taux de faux positifs pour différents groupes de la population. Est-ce que le taux de faux positifs est plus élevé pour des personnes habitant dans certains quartiers ? Construire un mécanisme de tests anti-discrimination peut se heurter aux principes du RGPD sur la non-utilisation des données sensibles.

Une autre condition imposée par la Cour concerne *l'intervention humaine*. Compte tenu de la présence inévitable d'erreurs dans ce type d'algorithme, la Cour impose pour chaque signalement positif un réexamen individuel par des moyens non-automatisés avant de procéder à des mesures de surveillance plus poussées[30]. L'intervention humaine est considérée comme une garantie importante pour les droits et libertés individuels[31]. Si le principe de l'intervention humaine ne fait pas débat, les modalités peuvent poser problème compte tenu des biais cognitifs humains. Les humains ont tendance à faire trop confiance aux recommandations algorithmiques, abandonnant leur sens critique[32]. Plusieurs techniques d'explication algorithmique permettent de réduire ces biais, mais le problème reste de taille. Le deuxième problème se pose spécifiquement pour le système de surveillance algorithmique mis en place par la loi du 24 juillet 2015. L'idée du législateur est que l'algorithme déclencherait des alertes sans que les enquêteurs humains puissent accéder aux données particulières de la personne. Ce serait seulement dans un deuxième temps, après autorisation du Premier ministre, que les enquêteurs pourraient accéder à l'information plus détaillée sur la personne et sur ses communications. Or, la Cour exige une intervention humaine avant la mise en place de mesures de surveillance supplémentaires donc avant l'autorisation complémentaire et l'identification de la personne, ce qui semble impossible puisque le décideur humain aura besoin de ces informations supplémentaires pour déterminer s'il s'agit d'un faux positif.

La dernière condition imposée par la Cour concerne *l'information des personnes*. L'autorité responsable de l'algorithme est tenue de publier des renseignements de nature générale sur le traitement, et dans l'hypothèse où l'autorité procède à une vérification plus ciblée, la personne doit être informée de manière individuelle. L'information doit intervenir dès que possible sans compromettre l'enquête en cours[33].

Pour résumer, la Cour impose sept conditions cumulatives qui devront trouver leur place dans tout système algorithmique créant un risque pour les droits et libertés individuels :

- Une adéquation entre le niveau d'ingérence causée par l'algorithme et l'objectif d'intérêt général poursuivi : le terrorisme peut justifier un niveau d'ingérence algorithmique élevé compte tenu de la gravité de la menace.
- Un contrôle institutionnel, soit par une juridiction, soit par une autorité indépendante, pour vérifier le respect de toutes les conditions.
- La transparence et l'explicabilité du modèle doit permettre un contrôle effectif de l'algorithme et de ses résultats.
- Le modèle doit démontrer sa fiabilité dans l'atteinte de l'objectif visé. Le niveau de faux positifs sera un critère clé dans cette analyse.
- Le modèle doit être non-discriminatoire, ce qui nécessite de pouvoir tester l'algorithme par rapport à différents groupes de la population.

25 CJUE 6 oct. 2020, aff. jtes C-511/18, C-512/18 et C-520/18, *La Quadrature du net (Assoc.)*, préc., pt 180.
26 *Ibid.*
27 *Ibid.*, pt 181 : « les modèles…ne sauraient être fondés sur ces *seules* données sensibles », ce qui laisse penser que le traitement de données sensibles serait quand-même possible (italiques fournies par l'auteur).
28 *Ibid.*
29 P. Bertail, D. Bounie, S. Clémençon et P. Waelbroeck, *Algorithmes : Biais, Discrimination et Équité*, 2019 hal-02077745.
30 CJUE 6 oct. 2020, aff. jtes C-511/18, C-512/18 et C-520/18, *La Quadrature du net (Assoc.)*, préc., pt 182.
31 V. not., art. 22(3) et consid. 71 du règl. 2016/679 (RGPD) ; Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle, Lignes directrices en matière d'éthique pour une IA digne de confiance, 8 avr. 2019, pts 64 et 65 ; A. Huq, A Right to a Human Decision, 106 Virginia L. Rev. 611 (2020).
32 Alberdi et al., Why are people's decisions sometimes worse with computer support ? Computer Safety, Reliability, and Security Proceedings, 5775, p. 18-31 (2009).
33 CJUE 6 oct. 2020, aff. jtes C-511/18, C-512/18 et C-520/18, *La Quadrature du net (Assoc.)*, préc., pt 191.

- Une intervention humaine sera nécessaire pour valider ou infirmer la recommandation algorithmique. Dans le cas examiné par la Cour, l'intervention humaine doit intervenir avant la prise de décision ayant des impacts défavorables pour l'individu. Dans d'autres contextes, l'intervention humaine pourrait intervenir *a posteriori*, notamment si le volume des décisions à traiter est très élevé [34]. L'intervention humaine suppose un jugement critique de la part du décisionnaire, ce qui peut être difficile en raison des biais cognitifs liés à l'automatisation.
- Enfin, l'information de la personne est nécessaire, de manière générale et de manière individuelle, lorsqu'une décision ciblant plus particulièrement la personne est prise.

III - QUELLE EST LA PORTÉE DE L'ARRÊT DE LA CJUE ?

Le champ d'application de l'arrêt de la CJUE pose question. Est-ce qu'il concerne uniquement l'interprétation de l'articles 15(1) de la directive 2002/58/CE dans le cadre particulier de la lutte contre le terrorisme, ou a-t-il une portée plus large ? Il existe deux raisons de penser que cet arrêt a une portée large. Premièrement, l'analyse de proportionnalité conduite par la Cour au titre de l'article 15(1) de la directive 2002/58/CE est très similaire sinon identique à l'analyse de proportionnalité conduite dans d'autres affaires concernant l'interprétation de la directive 95/46/CE et le RGPD [35]. Deuxièmement, dans son analyse, la Cour s'appuie sur la jurisprudence de la CEDH en matière de proportionnalité [36], et sur son avis de 2017 sur l'accord UE-Canada sur les données de passagers aériens (PNR) [37]. Or l'avis PNR de la CJUE ne concerne pas la directive 2002/58/CE mais s'appuie directement sur les dispositions de la Charte. Ainsi, les principes issus de l'arrêt du 6 octobre 2020 de la CJUE représentent selon nous une série de conditions incontournables d'une régulation des algorithmes conforme à la Charte. Ces principes rejoignent des principes que l'on retrouve dans le RGPD et dans les recommandations éthiques sur l'IA [38], mais ils ont l'avantage de s'appuyer directement sur la Charte. Ainsi, une question sur l'interprétation d'une disposition du RGPD par rapport à un traitement algorithmique conduirait probablement à la même liste de sept principes, même si l'intensité d'application de chaque principe sera variable en fonction des risques associés au traitement particulier.

Outre le contenu des sept conditions identifiées par la Cour, le point le plus important à retenir est qu'il ne suffit pas de décréter qu'un algorithme soit fiable et non-discriminatoire, car tout algorithme de prédiction comportera à la fois des erreurs (faux positifs, faux négatifs) et des biais. La question est plutôt le niveau acceptable de ces erreurs et biais. Cela nécessite la possibilité de mesurer le taux d'erreurs et de biais, et de fournir cette information à une autorité de contrôle pour que soit débattue la question de leur niveau. Le caractère excessif ou non du taux de faux positifs dépendra du contexte, de l'existence de mesures pour les réduire, et de l'effet de ces mesures sur d'autres paramètres de performance. Le paramétrage des algorithmes se réduit souvent à l'art du compromis, car chaque mesure de correction, par exemple une mesure pour réduire certaines discriminations, créera potentiellement d'autres biais ou discriminations ou dégradra la performance. Les autorités de contrôle devront rechercher une pondération équilibrée entre les performances algorithmiques et les effets indésirables sur les droits individuels.

34 On peut penser aux systèmes automatisés de retrait de contenus en ligne, par ex.
35 V. par ex., l'affaire CJCE, 16 déc. 2008, aff. C-73/07, *Satakunnan Markkinapörssi et Satamedia*, RSC 2009. 197, obs. L. Idot en ce qui concerne la dir. 95/46/CE et CJUE 24 sept. 2019, aff. C-136/17, *GC c/ Commission nationale de l'informatique et des libertés*, pt 58 AJDA 2019. 1839 ; *ibid*. 2291, chron. P. Bonneville, C. Gänser et S. Markarian ; D. 2020. 515, note T. Douville ; *ibid*. 2019. 2022, note J.-L. Sauron ; *ibid*. 2020. 1262, obs. W. Maxwell et C. Zolynski ; Dalloz IP/IT 2019. 631, obs. N. Martial-Braz ; Légipresse 2019. 515 et les obs. ; *ibid*. 687, étude N. Mallet-Poujol ; RTD eur. 2020. 316, obs. F. Benoît-Rohmer en ce qui concerne le RGPD.
36 CJUE 6 oct. 2020, aff. jtes C-511/18, C-512/18 et C-520/18, *La Quadrature du net (Assoc.)*, préc., pt 128.
37 Avis 1-/15 Accord PNR UE-Canada du 26 juill. 2017.
38 V. not. les lignes directrices précitées n. 31.