Quaderni
Communication, technologies, pouvoir

105 | Hiver 2021-2022
L'Intelligence Artificielle : raison et magie

Jusqu'où l'institution peut-elle être augmentée ? Pour une éthique publique de l'IA

How far can the institution be increased? For a public ethics of AI

Thierry Ménissier

Édition électronique
URL : https://journals.openedition.org/quaderni/2234
DOI : 10.4000/quaderni.2234
ISSN : 2105-2956

Éditeur
Les éditions de la Maison des sciences de l'Homme

Édition imprimée
Date de publication : 25 janvier 2022
Pagination : 73-88
ISSN : 0987-1381

Référence électronique
Thierry Ménissier, « Jusqu'où l'institution peut-elle être augmentée ? Pour une éthique publique de l'IA », *Quaderni* [En ligne], 105 | Hiver 2021-2022, mis en ligne le 02 janvier 2025, consulté le 08 janvier 2025. URL : http://journals.openedition.org/quaderni/2234 ; DOI : https://doi.org/10.4000/quaderni.2234

Jusqu'où l'institution peut-elle être augmentée ? Pour une éthique publique de l'IA

Thierry Ménissier

Institut de Philosophie de Grenoble — Université Grenoble Alpes

RÉSUMÉ

Cette contribution explore l'hypothèse que l'adoption par les services publics des outils technologiques contemporains informatiques et numériques, des algorithmes et des méga-données qui définissent l'intelligence artificielle (IA), induit une transformation importante du sens de l'institution. L'expertise des machines et la valeur du savoir des mathématiques-informatiques nourrissent la confiance envers l'IA, qui repose sur son efficacité pour assister les tâches d'administration et de gestion publiques. Mais elle entretient également une confusion entre la logique de la performance technique et la doctrine de l'État, à partir de la ressemblance entre des notions pourtant irréductibles les unes aux autres, par exemple entre fiabilité et confiance, entre généralité et universalité, entre objectivité et impartialité. Par le biais des services qu'ils rendent à l'institution et via l'influence qu'ils exercent déjà sur les comportements de ses agents et responsables, les systèmes d'IA acquièrent une action réelle qu'il est nécessaire de préciser et de conceptualiser. Cet article entend y contribuer à partir de l'idée qu'il s'agit d'une forme d'autorité, telle que la philosophie politique l'a interprétée.

ABSTRACT

How far can the institution be increased? For a public ethics of AI

This contribution explores the hypothesis that the adoption by public services of contemporary computer and digital technological tools, algorithms and mega-data that define artificial intelligence (AI), induces an important transformation of the meaning of the institution. The expertise of machines and the value of mathematical-compu tational knowledge feeds a form of confidence in AI, which is based on its undeniable effectiveness in assisting public administration and management tasks. But it also maintains a confusion between the logic of technical efficiency and the doctrine of the State, based on the similarity between notions that are irreducible to each other, for example between reliability and trust, between generality and universality, between objectivity and impartiality. Through the services they render to the institution and through the influence they already exert on the behaviors of its agents and managers, AI systems acquire a real action that needs to be specified and conceptualized. This article intends to contribute to this task, starting from the idea that it is a form of authority, as interpreted by political philosophy.

L'apparition contemporaine des technologies d'intelligence artificielle (IA) dans le cadre du fonctionnement des institutions publiques (la Défense, l'École, la Justice, le Gouvernement, etc.) suscite de l'intérêt pour une approche qui interroge les significations morales et politiques des progrès techniques. Et s'il apparaît stimulant de se pencher sur les relations entre les différentes activités auxquelles donnent lieu ces institutions et le recours à des outils informatiques et numériques, c'est que l'emploi actuel de ces derniers dans un secteur aussi important pour le fonctionnement des sociétés humaines conduit à faire l'hypothèse d'une possible transformation du sens des pratiques institutionnelles.

Les promesses recélées par le prochain déploiement massif des systèmes d'IA s'inscrivent dans une rhétorique bien huilée qui vante les avantages de l'innovation pour la bonne gestion privée et publique, en ravivant certains éléments de l'idéologie moderniste du progrès. Pourtant, le contexte contemporain apparaît également émotionnellement perturbé du fait de l'inquiétude suscitée par la puissance des nouvelles technologies, que traduisent par exemple des expressions chocs entremêlant les registres technique et politique telles que « datacratie » ou « algorithmocratrie[1] ». Nous voulons examiner l'hypothèse qu'au-delà de l'efficacité dans l'effectuation des tâches et des gains de commodité dans les usages, et derrière l'inquiétude affichée, il apparaît possible que se produise actuellement une transformation de fond à propos du sens des institutions. Cette contribution entend examiner cette hypothèse, et cela de deux manières différentes : d'une part, dans une perspective de philosophie politique et d'anthropologie de la technique à partir de la caractérisation des notions d'institution et de système sociotechnique, et de l'autre dans une optique qui conjugue psychologie sociale et éthique publique, en fonction de l'analyse des motifs pour lesquels – aussi iconoclaste qu'une telle suggestion puisse sembler – on peut affirmer que les technologies d'IA exercent d'ores et déjà une forme d'autorité.

LES INSTITUTIONS CIVILES ET LES USAGES DES DISPOSITIFS SOCIOTECHNIQUES

Dans les pays développés, les institutions ont depuis un certain temps recours aux technologies d'IA : depuis les années 2000, dans un contexte de débats et d'interrogations permanents sur le sens des transformations induites pour l'action publique, leurs gouvernements tirent bénéfice des nouvelles technologies de l'information[2]. Aujourd'hui, la gestion des

1. Voir par exemple *Pouvoirs*, n°164, 2018 : « La Datacratie » ; Olivier Derruine, « Algorithmocratie. L'économie numérique, un nouvel obscurantisme fondé sur la recherche de l'efficacité ? », *La Nouvelle Revue*, 2017/4 (n° 4), p. 26-32. Dans ces deux publications se laisse entendre l'inquiétude de la perte de la valeur de l'action de l'État en regard des méga-données, exploitables politiquement ou commercialement.
2. Voir D. West, *Digital Government : Technology and Public Sector Performance*, Princeton (New Jersey), Princeton University Press, 2005 ; J. Chevallier, « Vers l'État-plateforme ? », *Revue française d'administration publique*, 2018/3 (n°167), p. 627-637.

fonctions régaliennes de l'État ne se fait pas sans calculs algorithmiques, ni recours à des plateformes numériques délivrant et recueillant des informations, ni production de méga-données (big data). Quoiqu'une liste complète soit difficile à établir, on peut mentionner la défense du territoire et de la sûreté nationale, l'administration des populations par le biais de l'état civil et dans le cadre des politiques migratoires, la production et la gestion des ressources énergétiques ainsi que des richesses patrimoniales, la collecte des impôts et la régulation du budget, la gestion des flux d'élèves et d'étudiants dans les établissement d'enseignement, ou encore, dans les pays où le domaine de la Santé relève d'une tradition de gestion étatique, la gestion des données personnelles de santé.

Dans la transformation technologique des administrations et des services, certains secteurs de l'institution sont d'ailleurs susceptibles d'être plus exposés que d'autres. Un secteur particulier, celui de la Justice, peut ainsi valoir comme cas d'espèce pour guider la réflexion. L'apparition de la blockchain et des technologies d'IA (par exemple dans les Legal Tech) non seulement pour instruire les cas judiciaires mais également pour suggérer plus ou moins consciemment aux juges certaines décisions, et par conséquent pour inviter à rendre la justice d'une manière algorithmiquement assistée, offre depuis quelques années l'opportunité d'observer des évolutions intéressantes et de poser des questions importantes[3]. Ces avancées induisent à considérer le domaine judiciaire comme un véritable laboratoire de transformation de l'institution. À cet égard, les expérimentations rendues possibles par l'usage des big data en Justice suscitent de l'inquiétude liée à ce qui a été décrit comme la substitution, à une forme de langage – le langage alphabétique dans laquelle la loi s'est dite jusqu'à nos jours –, d'une autre forme, celle du calcul algorithmique et de la compilation de données[4]. Cette transformation provoque par elle-même des bouleversements qui font craindre que la « justice digitale » n'en soit pas vraiment une[5].

Nous voulons pour notre part suggérer l'idée que cette émergence technologique est susceptible de modifier le rapport que, de manière immémoriale, on a pris l'habitude d'établir entre l'institution et les services publics d'une part, et de l'autre ce que la société qui les a engendrés attend d'eux. Lorsque les algorithmes assurent les tâches jusqu'ici réalisées par les agents de l'État, que se passe-t-il pour l'institution ?

3. Voir notamment, parmi une littérature déjà abondante, B. Barraud, « Un algorithme capable de prédire les décisions des juges : vers une robotisation de la justice ? », *Les Cahiers de la Justice*, 2017/1 (n° 1), p. 121-139 ; P. De Filippi, A. Wright, *Blockchain and the Law. The Rule of Code*, Harvard, University Press, 2018 ; C. Dubois, F. Schoenaers, « Les algorithmes dans le droit : illusions et (r) évolutions », *Droit et Société*, 103/2019, p. 503-515.
4. Voir M. Hildebrandt, « Algorithmic Regulation and the Rule of Law », *Philosophical Transactions of the Royal Society*, vol. 376 issue 2128, 2018, accessible à l'URL https://royalsocietypublishing.org/doi/10.1098/rsta.2017.0355, consulté le 20/06/2021.
5. Pour une formulation précise de telles craintes à partir d'une problématisation approfondie de ces transformations et de l'hypothèse de la substitution d'un langage à l'autre, voir A. Garapon et J. Lassègue, *Justice digitale. Révolution graphique et rupture anthropologique*, Paris, PUF, 2018.

Peut-être que ces tâches sont aussi bien, voire plus efficacement effectuées que par le passé, mais le sens de l'institution ne s'en trouve-t-il pas modifié ?

« Lorsque les algorithmes assurent les tâches jusqu'ici réalisées par les agents de l'État, que se passe-t-il pour l'institution ? Peut-être que ces tâches sont aussi bien, voire plus efficacement effectuées que par le passé, mais le sens de l'institution ne s'en trouve-t-il pas modifié ? »

Pour répondre à ces questions, il importe de préciser cette notion d'institution. Avant d'être un concept que la sociologie puis l'ethnologie ont « élargi » dans leur enquête globale sur l'activité humaine[6], la notion d'institution, envisagée *lato sensu* en fonction de ses modalités civiles (i.e. non religieuses), renvoie au geste initial d'instituer, qui désigne la reconnaissance de la légitimité supérieure de certains pouvoirs du fait de leur cohérence, de leur pertinence sociale et de leur valeur éthique[7]. Le pouvoir ainsi reconnu offre un cadre explicite pour la mise en ordre et la stabilisation de la réalité sociale, pour la diffusion de valeurs, pour l'expression des différends et pour la résolution de conflits. Dans cette logique, au fil de sa construction philosophique dans la modernité, l'État de droit en est venu à constituer l'institution de référence capable de coordonner les activités humaines en tant qu'elles revendiquent, au-delà de leur dimension économique et sociale, d'avoir un sens éthique fondé sur l'intérêt général explicitement formulé[8]. La police et l'armée, la Justice (qui s'appuie également sur l'autorité intrinsèque de la loi et de l'idée du juste) et les académies savantes responsables de l'éducation et de la santé (qui quant à elles se fondent aussi sur la connaissance scientifique) s'adossent à cette institution de référence.

La transformation des cadres institutionnels, en fonction des changements sociaux, des réformes ou des révolutions politiques, ou encore des progrès scientifiques techniques engendre toujours certaines difficultés. Celles-ci se trouvent en partie induites par le régime temporel spécifique des institutions civiles. Elles n'ont en effet pas pour vocation fondamentale d'accompagner de tels changements, mais visent plutôt à garantir la stabilité sociale et la continuité temporelle les plus grandes possibles. Or

6. Voir J.-F. Kervégan, C. Schmidt et B. Zabel, « Les institutions et les exigences paradoxales de la modernité », *Trivium* [En ligne], 32/2021, p. 6, accessible à URL : http://journals.openÉdition.org/trivium/7496 , consulté le 24/05/2021.
7. Ce processus de légitimation a été analysé dans une perspective philosophique par J. Raz, notamment dans *The Authority of Law. Essais on Law and Morality*, Oxford, University Press, 1979.
8. La théorie rationnelle de l'État de droit regroupe dans une telle vue des auteurs aussi variés dans leurs options philosophiques particulières que Hobbes et Hegel. Le premier a fait reposer l'autorité souveraine de l'État sur le consentement des citoyens en redéfinissant la notion de liberté (voir T. Hobbes, *Léviathan. Traité de la matière, de la forme et du pouvoir de la république ecclésiastique et civile* [1651], chapitre XXI : « De la liberté des sujets », trad. de l'anglais par F. Tricaud, Paris, Sirey, 1983, p. 220-235) ; le second a consacré l'institution publique en considérant l'État comme « le rationnel en soi et pour soi », où « se compénètrent [...] l'universalité et la singularité » (G.W.F. Hegel, *Principes de la philosophie du droit* [1820], trad. J.-L. Vieillard-Baron, Paris, Flammarion, 1999, p. 299).

l'une et l'autre sont en étroite corrélation. Ainsi, lorsque Jean Bodin, dans *Les Six Livres de la République*, proposa sa définition de la souveraineté comme « *puissance absolue et perpétuelle d'une république*[9] », il soulignait le fait que peut être considérée comme souveraine la société politique qui choisit ses propres lois sans qu'on les lui impose, et qui, à partir de cette primauté, se trouve capable de structurer et de régenter son espace propre sans limitation de temps. Formuler un ordre juridique et coercitif incontestable et sempiternel, où force reste toujours à la puissance publique, telle fut la solution proposée par Bodin, solution pour laquelle les caractères classiques de la souveraineté sont la toute-puissance (à savoir, le droit légal et légitime de mettre hors d'état de nuire quiconque constituerait un danger pour l'État) et la perpétuité (le fait qu'il n'y ait, sur un plan de principe, pas de terme temporel à l'existence de l'ordre qui se revendique souverain)[10].

Toutefois, l'historicité des institutions, c'est-à-dire leur propre immersion dans un contexte historique toujours particulier, constitue elle-même un des ressorts de leur capacité à se transformer. Indissociables, les dimensions sociales et techniques concourent à ce type de modifications. Pour reprendre l'exemple de l'activité régalienne de la Justice, il faut par exemple considérer comme un fait historique la constitution de l'enceinte du Tribunal avec ses espaces fortement différenciés, et avec ses instruments spécifiques dont certains valent comme les symboles de sa légitimité via la désignation de certaines procédures judiciaires fondamentales. Par exemple, la barre où sont appelés les témoins est un dispositif technique permettant l'expression des témoins dans un procès et également un symbole du caractère public de cette expression. De tels dispositifs permettent de considérer le Tribunal comme le lieu du « rituel » de la ré-institution permanente de l'acte de juger de manière légale et légitime[11].

Ainsi envisagée, l'enceinte du Tribunal peut se comprendre comme un vaste dispositif sociotechnique. Ce qui signifie notamment que, à l'instar des autres secteurs d'activité de l'institution civile, la Justice a toujours intégré des éléments techniques, et que tout en consacrant une forme d'immuabilité dans ses principes fondamentaux, elle a accompagné, via l'intégration de tel ou tel dispositif particulier, le changement historique, et par suite accepté ce dernier[12]. Enfin, c'est aussi en tant que dispositif qu'elle est pleinement publique et légitime : le « Palais de Justice », parfaitement localisé dans chaque ville importante, représente même une des expressions irréductibles du langage de cette dernière. En d'autres termes, le caractère institutionnel du Tribunal trouve sa puissance de symbolisation non pas en dehors de la dimension socio-technique de ses procédures matérielles, mais bel et bien à travers elles.

9. Jean Bodin, *Les Six Livres de la république* [1576], Livre I, chapitre 8, Paris, Librairie Arthème Fayard, 1986, p. 179 sq.
10. Sur la relation entre le pouvoir absolu et la perpétuité chez Bodin, voir Julian Franklin, *Jean Bodin et la naissance de la théorie absolutiste* [1973], trad. J.-F. Spitz, Paris, PUF, 1993.
11. Voir A. Garapon, *Bien juger. Essai sur le rituel judiciaire*, Paris, Odile Jacob, 2001
12. Voir à ce propos P. Branco et L. Dumoulin, « La justice en trois dimensions : représentations, architectures et rituels », *Droits et société*, 2014/2 (n°87), p. 485-508.

Dans ce contexte, l'apparition des systèmes d'IA et de la blockchain – en dépit des différences entre ces dispositifs – fait légitimement craindre une transformation profonde de l'institution, voire fait poindre l'effroi de la dissolution pure et simple de sa forme traditionnelle. Surgit ce qui est perçu comme le risque de la « disruption[13]. » Mais ce que montre l'analyse du Tribunal en termes socio-techniques, c'est que la réalité des transformations en cours n'a rien d'évident, et que les éléments techniques nouveaux, lorsqu'on a correctement préparé leur intégration, favorisent à la fois la continuité du service public et la pérennité de l'institution. Comment envisager le sens profond de ces transformations pour mieux éclairer ce dilemme ?

« L'apparition des systèmes d'IA et de la blockchain — en dépit des différences entre ces dispositifs — fait légitimement craindre une transformation profonde de l'institution, voire fait poindre l'effroi de la dissolution pure et simple de sa forme traditionnelle. Surgit ce qui est perçu comme le risque de la *disruption*. »

L'IA POUR L'INSTITUTION : UNE ACTIVITÉ HUMAINE DANS LA DYNAMIQUE DE SES PROPRES OUTILS

Cette question apparaît complexe et gagnerait certes à être précisée en fonction des rôles, très variés, qu'on fait d'ores et déjà jouer aux algorithmes dans telle ou telle procédure judiciaire, de nombreux autres usages étant encore à inventer. Mais même envisagée de manière aussi générale, elle est intéressante à examiner, particulièrement si on l'appréhende à la lumière de la perspective de l'anthropologie historique des techniques. Cette discipline étudie la relation de l'humain à ses propres activités fondamentales (telles que se nourrir et se défendre, se déplacer et se soigner, acquérir et transmettre ses connaissances, etc.) en fonction des outils qu'il a mis à sa disposition à travers l'histoire. André Leroi-Gourhan, un de ses maîtres fondateurs, a proposé de relier l'observation méticuleuse des outils techniques (saisis à travers des « séries » qui sont comme des lignées d'êtres vivants) à l'étude des dispositifs sociaux ou culturels qui rendent efficace leur usage[14]. Il a également invité à mettre l'une et l'autre en perspective avec l'Évolution à laquelle, à l'instar des autres espèces animales, se trouve soumise l'humanité[15]. À travers ce prisme, il apparaît que les instruments techniques plus ou moins complexes inventés par l'humanité découlent de la station debout et du développement de certaines zones cérébrales particulières. Véritables orthèses cognitives produit par le dialogue de l'intelligence avec la matière, les objets et systèmes techniques sont donc

13. Ce terme de « disruption », de l'anglais « perturbation » ou « interruption », qualifie aujourd'hui dans la langue française les effets engendrés par les innovations mises en société tous azimuts, et produisant des effets de désorientation et de dépression, notamment pour les professionnels ou les agents de services publics qui subissent de tels changements. Voir B. Stiegler, *Dans la disruption : Comment ne pas devenir fou ?*, Paris, Les Liens qui Libèrent, 2016.
14. Voir A. Leroi-Gourhan, *Milieu et technique* (*Évolution et techniques*, tome II) [1945], Paris, Albin Michel, 1973 ; *Les Religions de la préhistoire*, Paris, PUF, 1964.
15. Voir A. Leroi-Gourhan, *Le Geste et la parole*, tome I : *Technique et langage* [1964], Paris, Albin Michel, 1995.

considérés comme les étapes d'un développement technique qui n'est pas encore achevé. Non seulement tout dispositif technique nouveau doit être appréhendé d'un point de vue historique et en regard de la fonction qu'il est appelé à effectuer dans les contextes sociaux et culturels qui sont les siens, mais encore l'histoire des techniques n'est pas achevée, et chaque étape enferme pour ainsi dire les humains dans l'étroit contexte d'usage dans lequel ils évoluent[16].

Ainsi, en poursuivant par hypothèses les analyses de Leroi-Gourhan, on pourrait formuler la suggestion que l'usage des technologies d'IA renvoie à une logique déterministe classique de la diffusion des innovations[17] et à une des options compatibles par l'activité même de l'institution. Comme tout dispositif sociotechnique, celle-ci vise à réaliser sa fonction de la manière la plus économique en effort possible ; en tant qu'institution, elle tend à inventer les modalités les plus rigoureuses du point de vue rationnel pour l'effectuation de ses tâches, à travers notamment l'établissement de l'objectivité et de l'impartialité qui permettent un traitement équitable de tous les citoyens. Or, les systèmes d'IA passent aujourd'hui pour de bons candidats afin de réaliser l'une et l'autre opérations. Dans une telle interprétation évolutionniste des techniques dont use l'institution, l'apparition de l'IA ainsi que la tentation d'un emploi varié et massif de cette dernière relèveraient des suites de ce que Leroi-Gourhan nomme « ***l'ascension prométhéenne***[18] » , lorsqu'il évoque les progrès de la métallurgie et ses effets sur les modes de vie collectifs et que nos contemporains nomment techno-déterministe dans la continuité du modèle cybernétique. Le développement des arts du feu a représenté pour les civilisations humaines une étape tellement importante qu'elle a déstabilisé les structures sociales et les a contraintes à se redéfinir. L'essor actuel de la puissance du calcul informatique ne constitue-t-elle pas une des conséquences de la même dynamique, étant donné que le « ***choix du feu*** » se poursuit aujourd'hui dans la « ***servitude électrique***[19] ? » Une telle vue offre des éléments de longue durée susceptibles de permettre d'appréhender des transformations aujourd'hui vécues de l'intérieur et sur le monde de la perturbation, étant entendu que toute innovation relève, dans sa conception comme dans sa mise en société, de certains rapports de forces, et qu'elle n'est jamais ni automatique dans

16. « *Imaginer qu'il n'y aura pas bientôt des machines dépassant le cerveau humain dans les opérations remises à la mémoire et au jugement rationnel, c'est reproduire la situation du Pithécanthrope qui aurait nié la possibilité du biface, de l'archer qui aurait ri des arquebuses, ou plus encore d'un rhapsode homérique rejetant l'écriture comme un procédé de mémorisation sans lendemain* », *Le Geste et la parole*, tome II, *La Mémoire et les rythmes*, Paris, Albin Michel, 1998, p. 75.
17. Voir à ce propos Pierre Doray, Jorge Niosi & Serge Proulx, *Diffusion de la technologie et des innovations*, Montréal, Presses de l'Université de Montréal, 2015.
18. *Le Geste et la parole*, tome I : *Technique et langage*, *op. cit.*, p. 245-246.
19. Pour reprendre les titres de volumes qui donnent à l'interprétation socio-anthropologique de la technologie contemporaine matière à critiquer les récents développements de l'informatique et du numérique : voir A. Gras, *Le Choix du feu. Aux origines de la crise climatique*, Paris, Fayard, 2007 ; G. Dubey et A. Gras, *La Servitude électrique. Du rêve de liberté à la prison numérique*, Paris, Éditions du Seuil, 2021.

son déploiement, ni évidente par ses seuls mérites techniques, ni légitime a priori ou de manière incontestable[20].

QUELLE FORME DE CONFIANCE ENVERS LES SYSTÈMES ALGORITHMIQUES ?

Si l'usage de systèmes d'intelligence artificielle pour la fonction publique apparaît non seulement humainement acceptable, mais aussi socialement tentante, c'est également en fonction du rôle accordé aux systèmes algorithmiques via la confiance qu'ils inspirent en termes d'expertise et d'objectivité. Cela conjugué à l'idéologie des bienfaits de l'automaticité technique pour régir tous azimuts l'activité humaine, et sans d'ailleurs qu'on dispose explicitement des repères éthiques pour décider jusqu'où peut et doit aller cette dernière, tout se passe alors comme si l'on conférait déjà une forme d'autorité aux machines.

Bien qu'un tel rapprochement entre l'usage des outils contemporains et la notion d'autorité puisse paraître iconoclaste, l'analyse de la relation de la confiance entretenue par les humains à l'égard des artefacts informatiques et numériques permet de rapprocher l'efficience des systèmes techniques d'IA et la dimension politique de la notion d'autorité. Car s'il semble contre-intuitif, le rapprochement n'en est pas moins nécessité par l'observation de la réalité. En effet, les systèmes algorithmiques et leur diverses expressions semblent, du fait de leurs performances alléguées et de leur fiabilité fortement médiatisée, capables de se substituer aux agents humains de l'État : aujourd'hui les pilotes des véhicules de combat des armées régulières et les forces de l'ordre déployées sur leurs théâtres d'opération, ou encore les fonctionnaires du Trésor Public et certains médecins de l'hôpital, demain sans doute les juges, la plupart des enseignants, et peut-être un jour certains membres du gouvernement.

On peut d'abord dresser le double constat que le déploiement contemporain de l'IA requiert certaines formes de confiance envers les machines, partagées aussi bien par leurs concepteurs que par leurs prescripteurs au sein de la sphère publique, et que ces formes de confiance apparaissent indispensables en tant que « ciment » des usages garantissant l'efficacité des machines. C'est dans un tel contexte que la question se pose de savoir s'il s'agit dans ces usages d'une réinvention authentique de la confiance ou d'une nouvelle forme de relation. La question se pose d'abord parce que la nature de la confiance s'avère philosophiquement d'elle-même peu claire[21]. Elle se pose ensuite parce que, selon l'approche traditionnelle qui considère la confiance comme un affect social, la notion n'en est pas spontanément employée pour traiter de la relation aux êtres inanimés. Dans une contribution importante pour la compréhension anthropologique de cet affect, le sociologue Louis Quéré formulait même certaines réserves :

20. Voir à ce propos D. Pestre (éd.), *Le gouvernement des technosciences. Gouverner le progrès et ses dégâts depuis 1945*, Paris, Éditions de la Découverte, 2014 ; T. Ménissier, *Innovations. Une enquête philosophique*, Paris, Hermann, 2021.
21. Voir G. Origgi, *Qu'est-ce que la confiance ?*, Paris, Librairie philosophique J. Vrin, 2008 ; M. Hunyadi, *Au début est la confiance*, Lormont, Éditions Le Bord de l'Eau, 2020.

« Il n'est pas sûr qu'il y ait un sens à parler de confiance lorsqu'il s'agit de se fier à la stabilité de l'environnement ou à la régularité de comportement de ses objets. En effet, le cas paradigmatique de la confiance est celui d'une relation de confiance entre deux personnes. Les traits caractéristiques de la confiance se maintiennent-ils hors de ce contexte ? Il est possible que nous commettions un abus de langage lorsque nous parlons de faire confiance à un objet ou à une institution[22]*. »*

L'auteur de ces lignes relayait un des présupposés souvent et depuis longtemps affirmé par les recherches en sciences sociales : la confiance constitue un affect socialement construit afin de conjurer la peur en s'assurant de la stabilité du lien social ainsi que de la validité du rapport entre l'investissement dans la relation avec autrui et l'espoir de certains gains en retour (non-agression, coopération)[23]. La confiance des usagers à l'égard des artifices techniques, bien que particulière, s'avère le plus souvent conçu à l'aune d'un comparable schéma[24]. Et si l'on admet qu'il est difficile de parler de confiance à l'égard d'un objet technique matériellement tangible et d'une institution dont on peut constater l'action concrète, a fortiori peut-on réellement parler de confiance lorsqu'on se fie à un ensemble de machines automatiques, éventuellement coordonnées entre elles et dont le fonctionnement global est opaque pour la grande majorité des usagers ?

« Peut-on réellement parler de confiance lorsqu'on se fie à un ensemble de machines automatiques, éventuellement coordonnées entre elles et dont le fonctionnement global est opaque pour la grande majorité des usagers ? »

Dans le même temps, il faut relever une certaine confusion sémantique, lorsque, dans notre actualité, des experts réunis par la puissance publique contribuent à alimenter l'idée d'une mise en relation nécessaire entre le système technique de l'IA et la notion de confiance. À cet égard, la thématique médiatique très puissante autour de l'IA « digne de confiance » (Trustworthy Artificial Intelligence) amplifie la confusion en faisant apparaître la confiance dont l'IA peut être investie comme un des nœuds de l'interrogation éthique d'aujourd'hui. En avril 2019, un groupe de 52 experts réunis par la Commission européenne avait publié ses « règles éthiques pour une IA de confiance[25]. » Leur rapport final indique plusieurs

22. L. Quéré, « La structure cognitive et normative de la confiance », *Réseaux*, 2001/4 (n°108), p. 125-152, p. 131.
23. Voir ces contributions classiques : D. Gambetta (eds), *Trust: Making and Breaking Cooperative Relations*, Oxford, Basil Blackwell, 1988 ; A. B. Seligman, *The Problem of Trust*, Princeton, University Press, 1997.
24. Voir Adriano Fabris, « Can We Trust Machines? The Role of Trust in Technological Environments », in Adriano Fabris (eds), *Trust. Studies in Applied Philosophy* (Epistemology and Rational Ethics, vol 54), Cham. Springer, 2020, p. 123-135.
25. Voir le rapport de la Commission Européenne, 2019, *Ethics Guidelines for Trustworthy AI*, accessible à l'URL https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top, consulté le 20/06/2021.

critères jugés nécessaires pour estimer cette confiance, tels que la supervision humaine, la transparence, la robustesse et la non-discrimination, etc. On remarque le caractère hétérogène de la liste fournie, hétérogénéité qui repose sur le mélange entre des critères purement techniques validant un usage fiable de la technologie basée sur la robustesse de cette dernière, et des critères plutôt éthiques basés sur les valeurs de la démocratie (telle que la non-discrimination). Il est de ce fait permis d'estimer que ce premier travail remis par les experts réunis par la Commission Européenne recouvre de si considérables enjeux de réassurance et de sécurisation psychologiques qu'à regarder les choses de près, la notion de confiance ne s'en trouve pas fondamentalement éclairée.

Plusieurs questions se posent en effet de manière plus précise à partir du moment où l'on restitue un certain nombre de nuances sémantiques importantes : d'abord, si la fiabilité avérée d'une IA représente à la fois la condition de possibilité de sa diffusion dans la société et le résultat de son usage efficace, est-elle assimilable à de la confiance au sens complet du terme ? Ensuite, l'assurance impliquée par la fiabilité de cette machine à travers ses usages, derrière la supposée réinvention de la confiance, ne risque-t-elle pas de créer la tentation d'abandonner à l'intelligence artificielle une partie des prérogatives humaines, avec d'autant plus de facilité que celle-ci paraît le soulager de ses rôles embarrassants (par exemple, administrateur, soldat et officier, juge et policier – autant de rôles à forte contrainte en termes de responsabilité) ? Et à partir de quel seuil cette tentation devient-elle dangereuse pour l'intégrité de la décision humaine ?

EXPERTISE AVÉRÉE, OBJECTIVITÉ PROJETÉE ET AUTORITÉ « POLITIQUE » DES MACHINES

Via les services qu'ils rendent et l'influence qu'ils exercent déjà sur les comportements, les systèmes d'IA exercent une action réelle qu'il convient de documenter et de conceptualiser. Si particulière qu'elle soit, il existe d'ores et déjà de la confiance envers l'action efficace des machines. Nous voulons à présent examiner l'hypothèse que cette confiance conditionne l'essor d'un nouveau genre d'autorité lié à l'efficience technologique.

La dynamique contemporaine d'évolution technique en informatique est notamment polarisée par l'essor de l'automatisation des machines, notamment par le biais de l'apprentissage automatique (machine learning). Des études ont montré que, alors qu'ils escomptent des effets favorables de cette tendance, les usagers manifestent paradoxalement à son propos une faible conscience ; et aussi que, second paradoxe, lorsqu'elle est consciente à leur esprit, elle apparaît assez peu acceptée[26]. Il y a bien un « enjeu de pouvoir » dans le rapport tissé entre efficience technique et idéologie promotrice de la technique[27] ; dit d'une manière plus générale, le pouvoir des algorithmes est « social[28]. » Pour qualifier la nature spéciale de

26. Voir par exemple T. Araujo, N. Helberger, S. Kruikemeier, C. H. de Vreese, 2020, "In AI we trust? Perceptions about automated decision-making by artificial intelligence", *AI & Society*, 35, 2020, p. 611-623.
27. L. Sfez, *Technique et idéologie. Un enjeu de pouvoir*, Paris, Éditions du Seuil, 2002.
28. D. Beer, « The social power of algorithms », *Information, Communication & Society*, 20/1, 2017, p. 1-13.

ce pouvoir, des conceptualités nouvelles et importantes ont déjà vu le jour[29]. Ces différents aspects attestent de la puissance de l'idéologie techniciste à l'œuvre dans les sociétés développées. L'angle de vue que nous adoptons pour notre part permet en outre de préciser en quoi ce pouvoir peut se voir légitimé au-delà de son efficience technologique. Nous voulons montrer que via l'implémentation des machines intelligentes au sein des institutions publiques, leurs concepteurs pourraient d'ores et déjà, ainsi que nous allons à présent l'établir, revendiquer pour elles une forme d'autorité.

La notion d'autorité, lorsqu'elle est définie comme le pouvoir humain le moins discutable ou le plus légitime, représente, dans la tradition éthique et politique moderne, un des principes de constitution de la subjectivité par elle-même. D'une part, dans la tradition philosophique, qu'elle soit idéaliste (Descartes) ou empiriste (Locke), le sujet qui revendique la maîtrise de lui-même (maîtrise de ses pensées, de son corps et de son travail) se pose comme autorité. De l'autre, dans la tradition éthique, la liberté repose sur l'assentiment volontaire à une institution ou aux commandements d'une personne, voire sur l'engagement en faveur de cette institution ou en appui de cette personne. En d'autres termes, l'autorité traduit un acte de la liberté humaine, et on pourrait dire qu'elle rend cette dernière manifeste[30]. Si l'on s'en tient à cette première approche, évoquer l'autorité des machines s'avère littéralement impossible, car l'autorité ainsi définie se fonde sur un principe d'autonomie compris dans la subjectivité, ce dont ne dispose à l'heure actuelle aucun système technique. Il est cependant permis de faire l'hypothèse que sur un autre plan il ne manque rien pour que l'on puisse évoquer à leur propos une forme d'autorité.

Le fait est que l'IA, via les flux de « méga-données » (Big data), contribue fortement à l'expertise qui appuie la décision stratégique des responsables des entités collectives (qu'elles soient publiques ou privées) et cela à tous les niveaux où elle est déployée. Aujourd'hui cette expertise se trouve toujours « augmentée » par l'IA, et cela, si l'on peut dire, exponentiellement, c'est-à-dire à un point tel que c'est la qualité de l'expertise (elle-même déterminée par la quantité des données) qui tend à valoir comme autorité. D'ailleurs, bien qu'elle prenne un tour manifeste dans l'essor récent de l'IA, la tendance à appuyer l'expertise sur les données n'est pas absolument nouvelle, puisqu'elle se fonde sur « la politique des grands nombres » qui constitue un trait distinctif de la façon dont la modernité a conçu l'exercice du pouvoir : sous l'effet du mouvement de fond de l'essor de la « raison statistique[31] », pour les institutions régaliennes, le rôle dévolu à l'expertise dans

29. Voir à cet égard la notion de « gouvernementalité algorithmique », dans T. Berns et A. Rouvroy, « Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation ? », *Réseaux*, 2013/1 n°177, p. 163-196.
30. Sur ces deux aspects, philosophique et éthique, de l'autorité, voir A. Kojève, *La notion de l'autorité* (1942), Paris, Gallimard, 2004 ; H. Arendt, « What is Authority ? » in *Between Past and Future. Eight Exercises in Political Thought* (1968), NYC, Penguin Books, 2006 ; R. Damien, *Éloge de l'autorité. Critique d'une (dé)raison politique*, Paris, Armand Colin, 2013.
31. Voir A. Desrosières, *La Politique des grands nombres. Histoire de la raison statistique*, Paris, Éditions de La Découverte, 2010 ; A. Supiot, *La gouvernance par les nombres*, Paris, Fayard, 2015.

le processus de décision devient prépondérant au point de valoir comme le ressort de la bonne décision. Tout se passe donc comme si s'esquissait, progressivement mais de manière catégorique, un démenti de la formule célèbre de Hobbes dans la version latine du *Léviathan,* formule pourtant fondatrice de l'autorité politique telle que le conventionnalisme moderne l'a conçue : « *Auctoritas non veritas facit legem*[32]. » Une des (fausses) évidences spontanées d'aujourd'hui suggère que c'est moins la force intrinsèque de la volonté que la qualité de l'expertise qui, dans la sphère publique, nourrit le principe de la décision légitime. De ce point de vue, l'essor de l'IA dans les pratiques d'expertise en vue de la décision ne fait que confirmer la dynamique moderne invisible et tacite à vouloir ou devoir « gouverner sans gouverner »[33].

Il serait même possible d'envisager une hypothèse selon laquelle le recours à l'IA s'inscrit dans un possible tacite fait de gouvernement sans experts et même sans expertise, la machine n'étant pas porteuse d'expertise mais d'une relation directe à la vérité (que le savoir expert traditionnel, issu de l'expérience, est censé permettre d'approcher – plus que de posséder).

« L'expertise des machines engage un processus de substitution de l'avis des conseillers humains par celui de ces experts artificiels et impersonnels que sont les algorithmes, qui conditionne l'émergence contemporaine d'une nouvelle source d'autorité légitime. »

Le paradoxe est que la logique même de la politique moderne porte en elle une telle tendance : c'est en restituant de manière précise la conceptualité propre à la naissance de l'autorité politique qu'il apparaît possible d'évoquer l'autorité des machines. Ainsi que l'expliquait encore Hobbes dans un chapitre important du *Léviathan*, l'autorité (Authority) d'une personne publique (Person) n'exprime rien d'autre que son statut et sa qualité d'auteur (Author) de ses propres actes, qui deviennent des décisions collectives du fait que les sujets le reconnaissent comme un acteur crédible (Actor)[34]. Doit être considéré comme une autorité, poursuivait l'auteur du *Léviathan,* l'être qui est crédible (en tout cas crédible à un degré suffisant) quant à sa capacité d'être l'auteur de ses propres actions mais également celui de celles d'autres personnes. Ce passage visait à déduire l'autorité politique d'une théorie de l'autorisation, qui constituait elle-même la pièce centrale de la construction de la notion abstraite d'une « personne civile » impersonnelle, considérée comme l'unique source valide de la légitimité publique[35]. Aujourd'hui, de manière comparable, l'expertise des machines,

32. « C'est l'autorité, et non la vérité, qui fait la loi (de l'État) », T. Hobbes, version latine du *Léviathan,* trad. F. Tricaud et M. Pécharman, Paris, Vrin & Dalloz, 2004, chapitre XXVI, 21, p. 210.
33. T. Berns, *Gouverner sans gouverner. Une archéologie politique de la statistique*, Paris, PUF, 2009.
34. Voir T. Hobbes, version anglaise du *Léviathan, Op. cit.*, p. 161-164.
35. Voir H. Warrender, *The Political Philosophy of Hobbes: his Theory of Obligation*, Oxford, Clarendon Press, 1957 ; L. Foisneau, *Hobbes. La vie inquiète*, Paris, Gallimard, 2016, notamment p. 68-93.

expression de la valeur d'objectivité prêtée aux mathématiques-informatiques, engage un processus de substitution de l'avis des conseillers humains par celui de ces experts artificiels et impersonnels que sont les algorithmes, qui conditionne l'émergence contemporaine d'une nouvelle source d'autorité légitime. Pourtant, le rôle accordé à l'informatique dans l'administration de l'institution n'est pas analysé comme il devrait l'être, à savoir, comme un changement de paradigme, la « vision numérique du monde » prenant la place autrefois dévolue à sa vision juridico-administrative[36]. Le grand Léviathan de l'État avait vu le jour sous l'effet d'un coup de force rationaliste opéré par Hobbes à l'encontre de la théologie médiévale[37]. Au fil de l'informatisation et de la numérisation progressive des services publics, ses ressorts intimes obéissent désormais à un « *Deus Ordinator* » nouvelle formule, celui de la raison computationnelle implicitement investie des pouvoirs d'ordonner le monde humain[38].

FORMULER UNE ÉTHIQUE PUBLIQUE DES USAGES DE L'IA

Fondée sur leur puissance de calcul et la haute qualité d'expertise permise par les méga-données, fondée également sur la réputation d'objectivité spontanément accordée aux artefacts créés par les mathématiques-informatiques dans la continuité du modèle cybernétique de Wiener, l'autorité potentiellement acquise par les systèmes algorithmiques n'est ni d'ordre philosophique, ni d'ordre éthique, mais d'ordre politique. En proposant cette interprétation de l'implémentation des technologies d'information contemporaines dans les institutions civiles, nous ne voulons nullement affirmer que, du fait de l'expertise des IA à qui responsables et usagers accordent leur confiance, il s'agit désormais de confier à des machines de calcul un pouvoir de type politique. Nous poussons seulement à son terme l'hypothèse selon laquelle l'analyse de la notion d'autorité en termes politiques, à l'aide de l'argumentation de Hobbes sur le processus de constitution de l'institution comme personne civile, révèle le rôle qu'implicitement on fait aujourd'hui jouer au système technique de l'IA. Selon nous ce rôle excède la simple dimension de l'assistance technique à la décision publique. Le calcul informatique, la notion d'information qui le caractérise, la fiabilité de l'expertise qu'il permet, la tendance de l'institution à se reconnaître dans l'idéologie de l'objectivité, tout cela concourt à ce que le système de l'IA soit accepté en tant que forme mentale considérée comme adéquate aux fonctions dévolues aux services publics et conforme aux finalités de l'institution.

Notre hypothèse porte à la fois sur les principes (à savoir, ceux qui sous-tendent la théorie de l'État) et sur les faits (qui concernent l'administration algorithmiquement assistée). Elle révèle l'ambiguïté de la situation contemporaine. D'une part, il existe une tentation de développer ces technologies

36. J. Lassègue, « L'Intelligence artificielle, technologie de la vision numérique du monde », *Les Cahiers de la Justice*, 2019/2 (n°2), p. 205-219.
37. Voir L. Foisneau, *Hobbes et la toute-puissance de Dieu*, Paris, PUF, 2001.
38. Sur la signification et les enjeux cognitifs, éthiques et politiques de la traduction du mot anglais « computer » par « ordinateur », proposée par le philologue Jacques Perret dans les années 1950, voir l'interprétation donnée par M. Alizart, *Informatique céleste*, Paris, PUF, 2017.

dans tous les domaines sensibles où la décision humaine doit assumer des responsabilités publiques importantes et de ce fait inquiétantes : par exemple, dans la Défense et la Sûreté, les politiques migratoires ou la Justice. De l'autre, un grand risque est encouru par les humains de se faire déposséder de cette prérogative, certes redoutable, mais fondatrice de la leur liberté : agir au nom de l'institution. Situation en regard de laquelle la formulation d'une éthique publique de l'IA, inspirée par un prométhéisme raisonnable, constitue une tâche importante et urgente.

Il est indéniable que, pour les institutions, les technologies de calcul sont aussi bien d'un usage extrêmement efficace que d'une utilité politique évidente. Les usages de l'IA participent d'un système de justification de l'utilité et de l'efficacité[39]. En effet, d'une part, elles agrègent de très nombreuses données de manière experte facilitant l'effectuation des services publics, et fournissent des prédictions souvent fiables, permettant de nourrir par des informations précises leur pilotage. La performance de l'État s'appuie normalement sur de tels outils. De l'autre, sous l'effet du discours de légitimation qui encadre les prouesses des mathématiques informatiques, il apparaît aisé de les faire passer pour plus objectives que les humains. La dimension fondamentale de l'impartialité de la chose publique se trouve par-là garantie. De telle sorte que face aux bienfaits des systèmes d'IA en termes d'efficacité, le citoyen peut se trouver pleinement satisfait en tant que citoyen : pour faire référence à ce qu'écrivait Hegel dans les *Principes de la philosophie du droit*, sa satisfaction vient de l'unité constatée de son intérêt personnel et de l'intérêt général[40]. En d'autres termes, en reprenant le cadre de la philosophie de l'État du philosophe d'Iéna, l'emploi de telles technologies souscrit pleinement au sens de l'institution pour un humain rationnel : la « satisfaction » du citoyen repose sur un double processus de reconnaissance : l'individu particulier reconnaît la légitimité de l'institution en constatant sa puissance du fait de l'efficacité des services publics, l'État est puissant et légitime à mesure de cette reconnaissance.

Le problème est que l'efficacité d'une institution n'est pas à chercher dans l'ordre uniquement pragmatique, mais doit être décelée dans un autre de faits, à savoir, dans une dimension symbolique qui, pour sa part, parce qu'elle relève du politique, se situe au-delà de la dimension technique et demeurait jusqu'à présent irréductible à cette dernière. Ainsi que l'écrivait également Hegel, « [...] *L'État n'est pas un mécanisme, mais constitue la vie rationnelle de la liberté consciente de soi, le système du monde éthique* [...][41]. » S'installe alors une forme de gêne, car à ce propos, il y a dans le recours contemporain aux systèmes d'IA comme un impensé, voire quelque chose de difficile à reconnaître, d'indicible, et qu'il serait tentant de vouloir dénier. La transformation en cours dans le rapport des humains à leurs institutions est potentiellement silencieuse, voire peut-être l'objet d'un déni collectif. La perte du sens politique de l'institution, sous l'effet de l'efficacité de services publics largement soumis à des systèmes

39. Voir Étienne Candel. « Les nouveaux outils du pouvoir. Tours et atours technologiques de l'autorité », *Quaderni*, n°99-100, Hiver 2019-2020, p.137-150.
40. G.W.F. Hegel, *Principes de la philosophie du droit*, § 261, *op. cit.*, p. 307-308.
41. *Ibidem*, § 270, p. 319.

de gestion administrative automatisés, constitue peut-être une tentation mortifère typique de notre temps. La rationalité technologique, expression de la modernité, n'a pas vocation à remplacer cette autre expression qu'est la rationalité politique symbolique ; mais, dans le fonctionnement de l'administration et de la gestion par l'État, si aujourd'hui elle l'assiste largement et l'augmente, elle peut facilement être tentée de se substituer à elle. Partout dans les États de droit, une éthique publique de l'IA se dessine certes dans le suivi des expérimentations d'implémentation des systèmes algorithmiques au sein des services publics, de même que dans la prudence qui préside à de telles expérimentations. Toutefois, ni l'une ni l'autre ne sauraient remplacer une réflexion de fond, par exemple organisée sur le mode du débat public d'intérêt national.