L'ÉVALUATION D'IMPACT ALGORITHMIQUE : UN OUTIL QUI DOIT ENCORE FAIRE SES PREUVES

Elisabeth LEHAGRE

Étude réalisée à la demande d'Etalab

Le présent rapport a été rédigé à la demande d'**Etalab** (Direction interministérielle du numérique – etalab.gouv.fr), dans le cadre de ses travaux sur les algorithmes publics.

Elisabeth Lehagre est chercheuse indépendante, spécialiste de la protection des données et de la régulation des technologies. Avocate de formation et juriste spécialisée en droit des nouvelles technologies, elle a travaillé pendant plus de 15 ans au sein d'entreprises internationales, notamment dans le secteur de l'IT avant d'évoluer vers une fonction de conseil en éthique du numérique et innovation responsable en créant la société Babotics. Elle est également enseignante à l'IMT Atlantique et l'EPITA en droit et éthique du numérique.

5 juillet 2021

SOMMAIRE

INTRODUCTION

Alors que l'utilisation de systèmes algorithmiques[1] est de plus en plus étendue, les effets de ces derniers sur les personnes et la société gagnent en visibilité. De nombreux travaux et rapports ont en effet souligné la diversité des enjeux liés au déploiement de ces systèmes, susceptibles d'influencer voire d'automatiser des décisions pouvant affecter négativement les individus, comme par exemple en matière d'autonomie ou d'équité[2]. De la sélection de contenus informationnels à la prédiction de récidive en matière de crime, l'utilisation des systèmes algorithmiques s'inscrit dans des environnements et contextes qui conditionnent leurs impacts à l'échelle tant individuelle que sociétale.

Face à cette « force d'impact » potentielle, l'appel à évaluer les effets de ces systèmes algorithmiques s'est fait de plus en plus pressant. Cet intérêt pour les questions d'impact s'est notamment précisé dans le cadre du recours à des systèmes algorithmiques par l'administration pour la prise de décisions administratives susceptibles d'affecter les personnes, de manière individuelle ou collective. Les réflexions menées à ce sujet sont éclairées par l'appréhension de la dimension particulière de l'action publique au sein d'une société démocratique et des principes de transparence et de redevabilité qui y sont généralement associés. C'est d'ailleurs dans la continuité de ces principes que les gouvernements du Canada et de la Nouvelle-Zélande ont proposé des outils d'évaluation d'impact algorithmique.

Dans le cadre de sa mission d'accompagnement des administrations dans la mise en œuvre du principe de transparence des algorithmes publics (la France étant l'un des rares pays à disposer d'un cadre juridique spécifique sur la transparence des algorithmes publics[3]), le département Etalab s'est intéressé aux questions liées à l'évaluation des impacts des algorithmes dans le secteur public et aux outils élaborés à cet effet.

Des propositions d'outils d'« évaluation d'impact algorithmique » ont en effet émergé, s'inscrivant dans la continuité des études et évaluations d'impact déjà existantes en matière de droits de l'homme, d'environnement ou de protection des données personnelles. Afin de mieux cerner cette tendance récente, il nous a paru intéressant d'examiner quelques-unes de ces propositions

---

[1] Selon la définition proposée dans la recommandation CM/Rec(2020) du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme, les systèmes algorithmiques sont « des applications qui, au moyen souvent de techniques d'optimisation mathématique, effectuent une ou plusieurs tâches comme la collecte, le regroupement, le nettoyage, le tri, la classification et la déduction de données, ainsi que la sélection, la hiérarchisation, la formulation de recommandations et la prise de décision ».

[2] Voir par exemple les rapports de la CNIL « Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle » (2017) et du Défenseur des Droits « Algorithmes : prévenir l'automatisation des discriminations » (2020).

[3] Art. L. 311-3-1, L. 312-1-3 et R. 311-3-1-2 du Code des relations entre le public et l'administration.

afin d'en appréhender plus précisément l'objet, la nature et l'utilisabilité dans le cadre de projets de développement et/ou de déploiement de systèmes algorithmiques.

Pour ce faire, nous avons sélectionné quelques exemples d'outils développés, entre 2018 et 2020, au niveau gouvernemental mais également par des organisations de la société civile. Le choix s'est porté sur ces outils en raison notamment de leur large diffusion et de leur caractère souvent « pionner » en tant que méthode et/ou outil ciblant plus précisément les systèmes algorithmiques et l'évaluation de leur impact sociétal. À ce jour, il apparait que les propositions résultent essentiellement d'initiatives anglo-saxonnes.

Outils d'évaluation examinés	
Société civile	**Gouvernement**
• Liste d'évaluation pour une IA de confiance par le groupe d'experts de haut-niveau en IA ou « GEHN IA » (constitué par la Commission européenne)[4] [Juillet 2020]	• « Évaluation de l'incidence algorithmique » - Secrétariat du Conseil du Trésor, Canada[5] [2018 - 2020]
• Rapport « *Algorithmic Impact Assessments* » de l'AI Now Institute (États-Unis)[6] [Avril 2018]	• « *Stakeholder Impact Assessment* » ou « SIA » - Government Digital Services, Office for Artificial Intelligence, Royaume-Uni[7] [2019]
• « *Ethics and Algorithms Toolkit* » proposé par un collectif de chercheurs et d'experts (États-Unis)[8] [2018]	• « *Algorithm Charter for Aotearoa New-Zealand* » - Stats NZ, Nouvelle-Zélande[9] [Juillet 2020]

En outre, des propositions et recommandations pour l'adoption de règlementations futures, intégrant cette dimension d'évaluation d'impact des systèmes algorithmiques, ont également été considérées.

[4] The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment (2020)
[5] https://ouvert.canada.ca/data/fr/dataset/5423054a-093c-4239-85be-fa0b36ae0b2e
[6] AI Now Institute: « Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability"; Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker (2018)
[7] https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety
[8] https://ethicstoolkit.ai/
[9] https://data.govt.nz/use-data/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter

Textes de prospective réglementaire examinés

- Recommandation CM/Rec (2020) du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme
- LIVRE BLANC de la Commission européenne - Intelligence artificielle - Une approche européenne axée sur l'excellence et la confiance (2020)
- *Algorithmic Accountability Act of 2019* (Proposition de loi, États-Unis)

Enfin, afin de compléter cette revue, plusieurs entretiens ont été réalisés, notamment avec des interlocutrices et interlocuteurs ayant participé à l'élaboration de ces outils et/ou suivant leur utilisation.

La présente étude opérée sur la base des outils et textes précités n'a pas vocation à revêtir un caractère exhaustif, du fait du nombre limité des propositions examinées, émanant, en outre, d'entités ou organisations basées dans des pays occidentaux. Toutefois, cette revue permet la réalisation d'un exercice préliminaire de compréhension qu'il nous a semblé important d'entreprendre car bien que plusieurs « évaluations d'impact algorithmique » soient d'ores et déjà proposées, il n'en ressort pas avec évidence une « ligne claire » quant à leur définition et degré d'utilisabilité pratique. C'est pourquoi nous tenterons de proposer, dans un premier temps, une définition de la notion d'« évaluation d'impact algorithmique » au regard de sa finalité générale (I) avant d'en explorer les formes et les contenus possibles (II) ainsi que leur utilisabilité, notamment dans le secteur public (III).

I – QU'EST-CE QU'UNE ÉVALUATION D'IMPACT ALGORITHMIQUE ?

La notion d'« évaluation d'impact algorithmique » peut être associée à des « réalités d'évaluation » différentes, que ce soit au regard de la nature des systèmes et des effets à évaluer et/ou des méthodes d'évaluation. Cela invite dès lors à considérer une définition générique (A) permettant de viser un outil d'auto-évaluation protéiforme (B).

A. Une définition générique pour une notion aux dimensions multiples

La notion d'« évaluation d'impact algorithmique » découle d'une traduction littérale de la formulation anglaise « *Algorithmic Impact Assessment* », largement employée dans la plupart des travaux consacrés à ce sujet[10]. L'évaluation d'impact algorithmique s'inscrit dans la continuité d'une démarche d'« *impact assessment* », initiée dans le monde anglo-saxon et consistant en une pratique d'évaluation *ex ante* fondée sur un calcul coûts – avantage, utilisée comme outil d'aide à la décision. Cette approche est aujourd'hui déclinée dans différents domaines afin d'évaluer les effets d'un projet ou d'une norme sur, par exemple, l'environnement, les droits de l'homme ou la protection des données personnelles. Les formulations anglophones retenues sont assez homogènes pour désigner ces évaluations : « *Environmental Impact Assessment* » en matière environnementale, « *Human Rights Impact Assessment* » en matière de droits de l'homme ou « *Data Protection Impact Assessment* » pour la protection des données.

Les termes français utilisés présentent un caractère plus aléatoire allant de l'évaluation (« évaluation environnementale ») à l'étude (« étude d'impact sur les droits de l'homme »), voire à l'analyse d'impact (« analyse d'impact relative à la protection des données »). Une distinction pratique semble cependant s'opérer entre l'« **évaluation »** qui tend à désigner un processus d'évaluation continue[11] ou *a posteriori*[12] et l « **étude** » qui vise plutôt l'évaluation d'impact *ex ante*[13].

Aux fins de simplicité et de lisibilité du présent document, le choix de la formulation « **évaluation d'impact algorithmique** » (ou « **EIA** ») a été retenue en ce

---

10 Voir par exemple : Moss, Emanuel and Watkins, Elizabeth and Metcalf, Jacob and Elish, Madeleine Clare, *Governing with Algorithmic Impact Assessments: Six Observations* (April 24, 2020) ; Kaminski, Margot E. and Malgieri, Gianclaudio, *Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations* (September 18, 2019). International Data Privacy Law, 2020, forthcoming., U of Colorado Law Legal Studies Research Paper No. 19-28.

11 C'est le cas par exemple de l'évaluation environnementale qui désigne le processus permettant d'estimer l'incidence d'un projet ou programme sur l'environnement et dont l'étude d'impact en constitue la base de départ préalable (art. L.122-1 et s. du Code de l'environnement).

12 Décret n°98-1048 du 18 novembre 1998 relatif à l'évaluation des politiques publiques.

13 Art. 8 de loi organique n° 2009-403 du 15 avril 2009 relative à l'application des articles 34-1, 20 et 44 de la Constitution (étude d'impact des projets de loi).

qu'elle permet de couvrir les différentes caractéristiques des EIA examinées, susceptibles de s'appliquer tout au long du cycle de vie des systèmes algorithmiques.

Car ce sont bien ces systèmes qui sont ici l'objet de l'évaluation afin d'en identifier les possibles effets. En cela, l'évaluation d'impact algorithmique diffère sémantiquement des autres types d'évaluation et d'étude d'impact en ce qu'elle s'attache aux effets des systèmes algorithmiques et non aux effets qu'un projet ou une norme pourrait avoir sur eux[14].

Comme nous le verrons, les objets et domaines d'évaluation peuvent varier d'une EIA à l'autre. Toutefois, il semble que toutes soient destinées à procéder à une évaluation intégrant une dimension sociétale avec pour visée, le bien-être tant individuel que collectif. L'objectif est ainsi d'identifier les impacts susceptibles d'affecter négativement les personnes à cet égard, pour tenter d'en limiter les effets. C'est ce qui nous conduit, à ce stade et avec pour objectif d'englober les différentes réalités de ces outils, à adopter une définition générique de l'évaluation d'impact algorithmique comme « **outil et/ou processus destiné(s) à évaluer, limiter et suivre les effets d'un système algorithmique pouvant affecter négativement les personnes et/ou la société, tout au long de son cycle de vie** ».

Cette définition constitue une proposition parmi d'autres plus spécifiques[15]. Certes perfectible, elle permet néanmoins de désigner la fonction générale d'une EIA qui peut se présenter sous différentes formes tout en restant basée sur un exercice d'auto-évaluation.

B. Un outil d'auto-évaluation aux formes variées : du guide de bonnes pratiques à l'outil de calcul de risque automatisé

Un élément commun aux différentes propositions d'EIA examinées peut d'ores et déjà être identifié. Il s'agit du recours à l'auto-évaluation comme base de l'identification des impacts associés aux systèmes algorithmiques (i). Toutefois, les modalités de cette approche commune varient selon les EIA qui présentent des formes et contenus variés (ii).

---

14 Ada Lovelace Institute "Examining the Black Box: Tools for assessing algorithmic systems" (2020), Endnote 2, p. 24.
15 Ada Lovelace Institute "Examining the Black Box: Tools for assessing algorithmic systems" (2020).

(i) Une approche commune : l'auto-évaluation

L'auto-évaluation s'inscrit dans le cadre d'une **approche fondée sur la gestion des risques** (« *risk-based approach* ») comme mode d'autorégulation non prescriptif et probabiliste. La réalisation d'objectifs s'envisage au regard des niveaux de risque et des mesures pouvant être mises en œuvre pour les limiter. Selon cette même approche, l'évaluation d'impact algorithmique se présente comme un outil d'aide à la décision basé sur un « calcul » de risques au regard des effets potentiellement négatifs de l'utilisation de systèmes algorithmiques sur les personnes. Cette évaluation du risque est réalisée au regard d'objectifs (généralement d'utilisation des systèmes) et a essentiellement vocation à permettre d'identifier quelles mesures peuvent être mises en place pour limiter voire supprimer les risques.

L'auto-évaluation marque également une grande autonomie des acteurs qui souhaitent procéder au développement d'un projet ou, pour ce qui est des EIA, au déploiement d'un système algorithmique. Cette approche peut poser la question de la qualité de juge et partie d'une entité qui a pour objectif de développer un système et qui se charge également d'en évaluer les effets. Nous verrons que cet aspect emporte quelques interrogations en matière d'utilisabilité et d'opposabilité d'une EIA[16].

L'auto-évaluation n'en demeure pas moins un élément qui apparaît structurellement attaché à l'ensemble des EIA examinées même si la forme de ces outils diffère.

(ii) Une auto-évaluation protéiforme

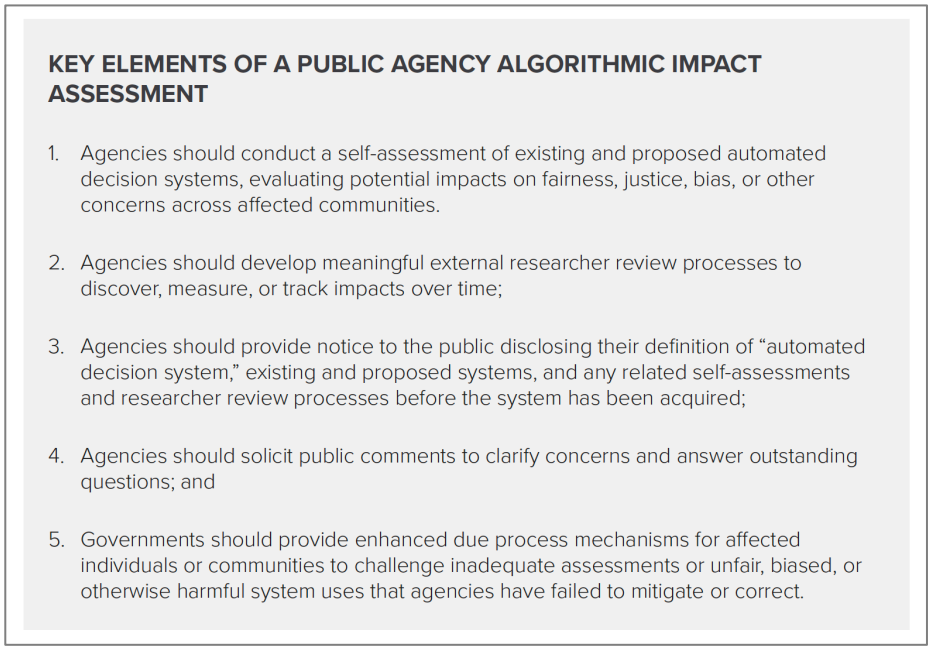
Les propositions d'EIA montrent en effet une variété d'approches aux formes multiples, allant du guide de bonnes pratiques à un outil en ligne de calcul automatique du risque.

- **Le cadre de bonnes pratiques**

L'EIA peut en effet se présenter sous la forme d'un **ensemble de bonnes pratiques**. C'est cette forme que retient l'AI Now Institute dans son rapport « *Algorithmic Impact Assessments : A Practical Framework for Public Agency Accountability* ». Cinq recommandations pratiques essentielles y sont proposées, parmi lesquelles se trouvent l'invitation à réaliser une auto-évaluation d'impact, la mise en place de procédures pour la vérification et le suivi externe des impacts

[16] Cf. *infra* p. 24

identifiés et la mise en œuvre de mesures de transparence et de consultation publique.



KEY ELEMENTS OF A PUBLIC AGENCY ALGORITHMIC IMPACT ASSESSMENT

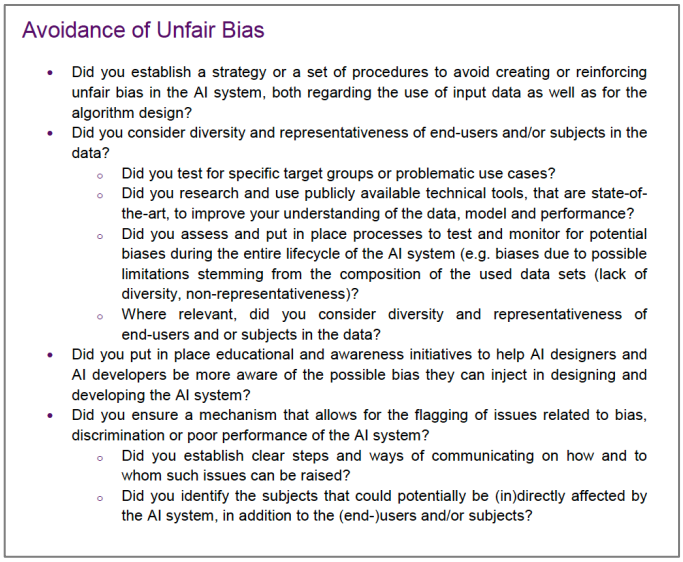
1. Agencies should conduct a self-assessment of existing and proposed automated decision systems, evaluating potential impacts on fairness, justice, bias, or other concerns across affected communities.
2. Agencies should develop meaningful external researcher review processes to discover, measure, or track impacts over time;
3. Agencies should provide notice to the public disclosing their definition of "automated decision system," existing and proposed systems, and any related self-assessments and researcher review processes before the system has been acquired;
4. Agencies should solicit public comments to clarify concerns and answer outstanding questions; and
5. Governments should provide enhanced due process mechanisms for affected individuals or communities to challenge inadequate assessments or unfair, biased, or otherwise harmful system uses that agencies have failed to mitigate or correct.

Extrait du rapport de l'AI Now Institute "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability" (2018)

Ces recommandations d'ordre général sont destinées à proposer un cadre pouvant guider les organisations gouvernementales qui souhaiteraient développer des pratiques d'évaluation d'impact des systèmes de décision automatisée. Elles ne visent pas de questions spécifiques à traiter à la différence d'autres outils qui vont plus précisément identifier des questions à se poser pour évaluer les impacts des systèmes algorithmiques.

- **La liste de questions**

Certains outils d'EIA vont privilégier une approche basée sur l'identification de questions à considérer pour aider à l'évaluation d'impact des systèmes algorithmiques. La nature et l'objet de ces questions peuvent varier[17]. Cette forme est notamment reprise dans l'EIA proposée par le groupe d'experts de haut-niveau en IA constitué par la Commission européenne.



Avoidance of Unfair Bias

- Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design?
- Did you consider diversity and representativeness of end-users and/or subjects in the data?
  - Did you test for specific target groups or problematic use cases?
  - Did you research and use publicly available technical tools, that are state-of-the-art, to improve your understanding of the data, model and performance?
  - Did you assess and put in place processes to test and monitor for potential biases during the entire lifecycle of the AI system (e.g. biases due to possible limitations stemming from the composition of the used data sets (lack of diversity, non-representativeness)?
  - Where relevant, did you consider diversity and representativeness of end-users and or subjects in the data?
- Did you put in place educational and awareness initiatives to help AI designers and AI developers be more aware of the possible bias they can inject in designing and developing the AI system?
- Did you ensure a mechanism that allows for the flagging of issues related to bias, discrimination or poor performance of the AI system?
  - Did you establish clear steps and ways of communicating on how and to whom such issues can be raised?
  - Did you identify the subjects that could potentially be (in)directly affected by the AI system, in addition to the (end-)users and/or subjects?

Extrait de la liste d'évaluation pour une IA de confiance par le groupe d'experts de haut-niveau en IA

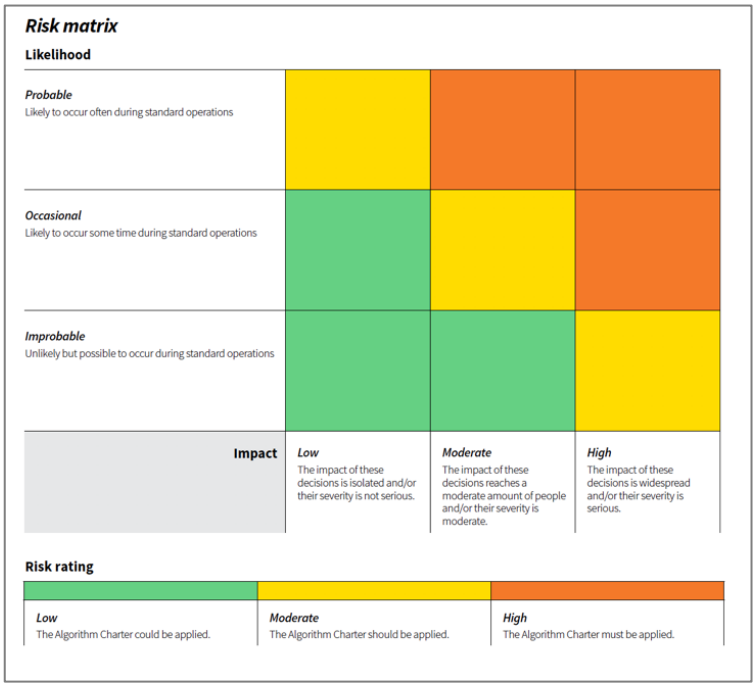
[17] Cf. *infra* p. 17 et s.

L'EIA invite ainsi à s'interroger, faisant de cet outil un support privilégié de réflexion et de discussion via un travail d'identification des impacts potentiels des systèmes algorithmiques. Certaines EIA vont également s'attacher à cibler plus précisément l'évaluation des risques attachés à ces impacts.

- **La matrice des risques**

Des matrices des risques sont parfois proposées afin d'identifier un niveau de risque associé au projet de déploiement d'un système algorithmique, sur la base de questions et/ou d'indicateurs prédéterminés.

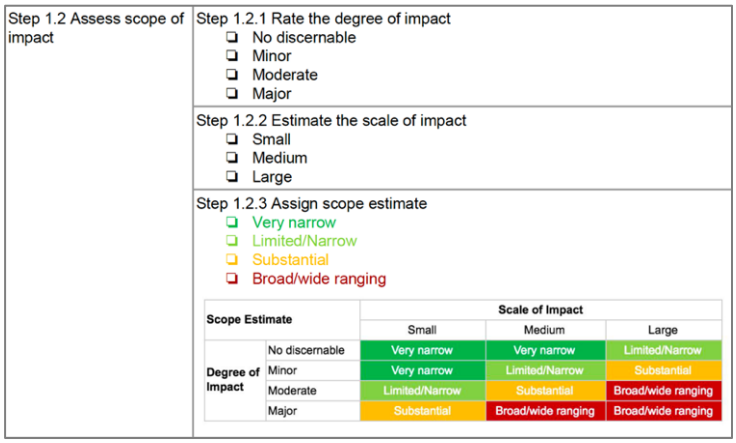
Le gouvernement de Nouvelle-Zélande a ainsi intégré une matrice des risques dans sa Charte sur les Algorithmes. Cette matrice vise à quantifier la probabilité d'effets négatifs indésirables sur le bien-être des personnes (improbable, occasionnel, probable) par rapport à son degré d'impact au regard du nombre de personnes potentiellement affectées (faible, modéré, élevé) pour aboutir à un niveau de risque qui pourra être faible, modéré ou élevé.



Matrice des risques proposée dans la Charte sur les Algorithmes (Nouvelle-Zélande)

Dans cet exemple, la matrice des risques apparaît comme un support d'**auto-évaluation ouverte**, sans questions spécifiques et laissant une grande autonomie aux personnes réalisant l'EIA pour quantifier l'impact et le risque. D'autres outils vont préférer guider plus précisément l'évaluation en intégrant une matrice des risques dans une approche d'**auto-évaluation plus fermée.** C'est notamment l'approche retenue dans l'EIA « *Ethics and Algorithms Toolkit* » qui propose une évaluation basée sur un ensemble de questions fermées et de

matrices de risques associées, permettant d'attribuer des niveaux de risques prédéterminés, applicables à des domaines d'impact pré-identifiés. Par exemple, si un impact en matière de liberté (« *Liberty/Freedom* ») est identifié pour le système algorithmique considéré, il est ensuite demandé de cocher les cases qui correspondent au degré, à l'échelle et à l'étendue de l'impact (de faible à substantiel) pour permettre une identification du niveau de risque via la matrice proposée dans l'outil.



Step 1.2 Assess scope of impact

Step 1.2.1 Rate the degree of impact
- ❑ No discernable
- ❑ Minor
- ❑ Moderate
- ❑ Major

Step 1.2.2 Estimate the scale of impact
- ❑ Small
- ❑ Medium
- ❑ Large

Step 1.2.3 Assign scope estimate
- ❑ Very narrow
- ❑ Limited/Narrow
- ❑ Substantial
- ❑ Broad/wide ranging

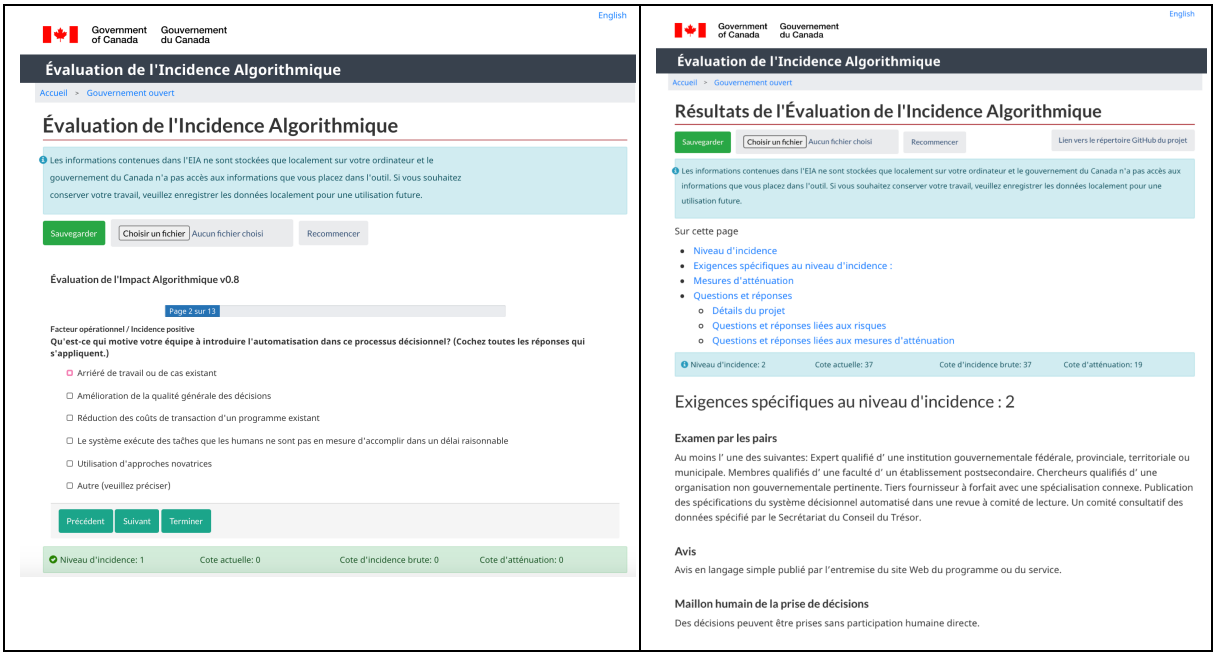
Scope Estimate		Scale of Impact		
		Small	Medium	Large
Degree of Impact	No discernable	Very narrow	Very narrow	Limited/Narrow
	Minor	Very narrow	Limited/Narrow	Substantial
	Moderate	Limited/Narrow	Substantial	Broad/wide ranging
	Major	Substantial	Broad/wide ranging	Broad/wide ranging

Extrait de l'outil "Ethics and Algorithms Toolkit" Worksheet for Part 1

On observe ainsi que ce type d'EIA associe clairement l'impact à un niveau de risque qui peut être évalué voire calculé automatiquement comme cela est le cas avec l'EIA proposée au Canada.

- **Le calcul automatisé de risque**

Parmi les EIA examinées, l'outil proposé par le Conseil du Trésor du Canada se démarque en proposant un **outil en ligne calculant automatiquement le niveau de risque** ou d'incidence algorithmique d'un système, sur la base d'un questionnaire à compléter. En fonction des réponses apportées aux questions, l'outil va calculer un niveau d'incidence pour le système concerné allant du niveau I (faible incidence) au niveau (IV) (incidence très élevée) qui conduira à la formulation de recommandations particulières.



Extrait d'une simulation d'EIA (outil en ligne proposé par la Conseil du Trésor du Canada)

Cet outil témoigne de la diversité des EIA qui, selon les formes adoptées, semblent osciller entre outils de sensibilisation axés sur l'identification des impacts et outils d'évaluation du risque. L'évaluation des risques et des impacts se mêlent et se confondent dans des EIA aux contenus eux-aussi variés. Mais que contient exactement une EIA ? Existe-t-il une « recette » type, reprenant une liste d'éléments à évaluer ?

II – QUE CONTIENT UNE ÉVALUATION D'IMPACT ALGORITHMIQUE ?

À des fins d'utilisabilité d'une EIA, il parait indispensable d'identifier ce sur quoi va porter l'évaluation et quels sont les éléments à considérer pour réaliser cette évaluation. Au regard des EIA examinées, il apparaît que leur contenu peut varier d'une proposition à l'autre que ce soit sur l'objet et les domaines d'impact à évaluer (A) ou concernant les modalités d'évaluation (B).

A. L'identification de l'objet et des effets à évaluer : sur quoi porte l'évaluation ?

(i) L'« objet algorithmique »

L'objet même de l'EIA est en effet diversement défini, allant des « systèmes d'IA », aux « systèmes de décision automatisée » en passant par les « algorithmes ». Ces notions renvoient à un champ d'application de l'évaluation plus ou moins

large et une vision plus ou moins ciblée des applications de systèmes algorithmiques les plus « critiques ».

En faisant le choix d'exiger ou de recommander une EIA pour les « **systèmes décisionnels automatisés** », le Conseil du Trésor du Canada (tout comme l'AI Now Institute qui vise les « systèmes de décision automatisée ») associe la réalisation d'une EIA à un type de systèmes algorithmiques. L'objectif est ainsi de « cibler les efforts » sur des systèmes considérés comme plus susceptibles d'impacter négativement et de manière significative les personnes. Ici, c'est la finalité d'usage des systèmes algorithmiques, en tant qu'outils d'aide à la décision humaine voir de décision automatisée directe qui permet d'identifier les systèmes les plus critiques et pour lesquels une évaluation d'impact serait nécessaire. Il s'agit en quelque sorte d'une présélection opérée sur la base de l'anticipation d'impacts plus conséquents de ce type de systèmes dès lors qu'ils permettent d'influencer ou de remplacer la décision humaine.

Cette position n'est pas reprise dans les autres outils examinés qui privilégient un objet d'évaluation plus large comme les « **systèmes d'IA** » ou les « **algorithmes** » sans en donner une définition précise. L'objectif assumé est ici de ne pas limiter l'évaluation d'impact au regard de la fonctionnalité particulière d'un système algorithmique. Cette approche envisage la réalisation d'une EIA sur tout type de systèmes d'IA ou d'algorithmes pour en évaluer les risques et impacts au regard de leurs contextes d'utilisation.

L'objet de l'évaluation peut ainsi varier même si les différentes propositions relèvent toutes de la sphère « algorithmique ». Il n'existe pas à ce jour d'arbitrage sur l'objet « type » d'évaluation à retenir, lequel arbitrage s'opérera peut-être du fait de l'usage de ces EIA et de leur degré d'utilisabilité. En effet, à ces objets plus ou moins identifiés s'ajoutent la question de leur compréhension, nécessaire à la mise en œuvre effective (et accessible) d'une EIA[18].

En outre, il n'y a pas que l'objet de l'évaluation à considérer. Pour réaliser une évaluation des effets des systèmes algorithmiques, encore faut-il savoir quels sont les effets à mesurer et surtout, au regard de quoi.

(ii) Les « domaines d'impact »

Les EIA montrent une variété des « domaines d'impact » à considérer pour évaluer les effets d'un système algorithmique, pouvant notamment revêtir une dimension sociétale et/ou juridique.

[18] Cf. *infra* p. 22

- **Impacts sociaux**

Les EIA examinées ont pour point de départ commun le besoin d'évaluer des impacts négatifs et indésirables que le déploiement de systèmes algorithmiques peut avoir sur les personnes et la société. Le domaine global d'impact envisagé est ici social. Il est possible d'y associer la définition donnée par le Conseil supérieur de l'économie sociale et solidaire qui désigne l'impact social comme « l'ensemble des conséquences (évolutions, inflexions, changements, ruptures) des activités d'une organisation tant sur ses parties prenantes externes (bénéficiaires, usagers, clients) directes ou indirectes de son territoire et internes (salariés, bénévoles, volontaires), que sur la société en général »[19].

Une telle définition, appliquée aux effets des systèmes algorithmiques, ne figure cependant pas dans les EIA qui, pour certaines, évoquent la dimension sociétale de ces effets sans apporter plus de précisions. Des thématiques propres à ce domaine d'impact sont toutefois mentionnées, invitant à évaluer les effets du déploiement des systèmes sur le **bien-être**, la **santé**, l'**environnement** ou l'**équité**.

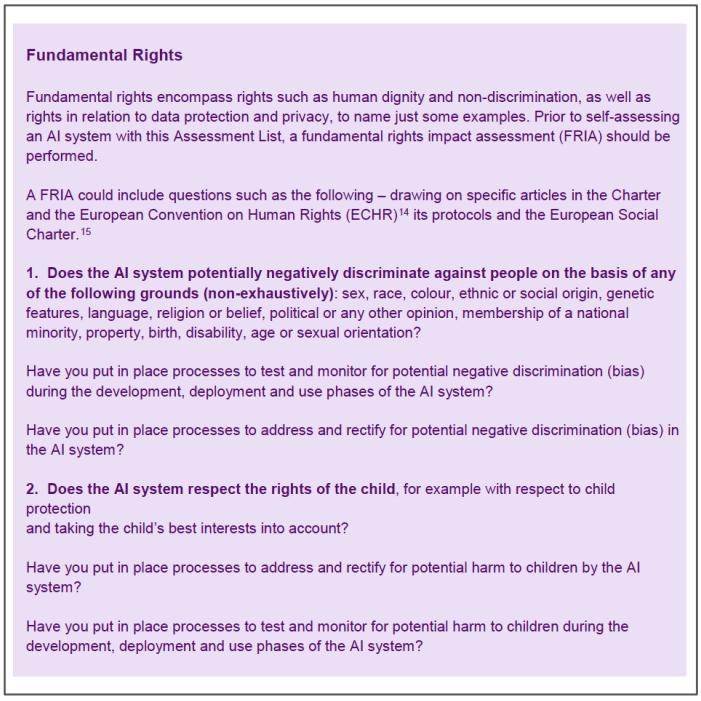
Au titre du « domaine d'impact social », les EIA évoquent également régulièrement les effets sur la **vie privée** ou en matière de **discrimination**, témoignant d'un recoupement assez symptomatique de la confusion entretenue entre impacts sociaux et impacts sur les droits des personnes et notamment les droits fondamentaux.

- **Impacts sur les droits fondamentaux**

Certains des outils examinés intègrent en effet les droits fondamentaux dans les domaines d'impacts à évaluer. Ces derniers peuvent être évoqués de manière générale (comme par exemple dans l'outil proposé par le Conseil du Trésor du Canada qui fait référence aux « droits ou libertés des personnes ») ou de manière plus spécifique en visant des droits particuliers comme le droit à la vie privée ou à la non-discrimination.

[19] Groupe de travail du CSESS sur la mesure de l'impact social « La mesure de l'impact social : Après le temps des discours, voici venu le temps de l'action » 2011.

L'EIA proposée par le GEHN IA va plus loin en prévoyant deux niveaux d'évaluations, le premier visant la réalisation d'une **étude d'impact sur les droits fondamentaux**[20] avant de réaliser, dans un second temps, une auto-évaluation basée sur des questions plus spécifiques et techniques liées à l'utilisation des données et aux fonctionnalités des systèmes d'IA.



Fundamental Rights

Fundamental rights encompass rights such as human dignity and non-discrimination, as well as rights in relation to data protection and privacy, to name just some examples. Prior to self-assessing an AI system with this Assessment List, a fundamental rights impact assessment (FRIA) should be performed.

A FRIA could include questions such as the following – drawing on specific articles in the Charter and the European Convention on Human Rights (ECHR)[14] its protocols and the European Social Charter.[15]

1. **Does the AI system potentially negatively discriminate against people on the basis of any of the following grounds (non-exhaustively)**: sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation?

Have you put in place processes to test and monitor for potential negative discrimination (bias) during the development, deployment and use phases of the AI system?

Have you put in place processes to address and rectify for potential negative discrimination (bias) in the AI system?

2. **Does the AI system respect the rights of the child**, for example with respect to child protection
and taking the child's best interests into account?

Have you put in place processes to address and rectify for potential harm to children by the AI system?

Have you put in place processes to test and monitor for potential harm to children during the development, deployment and use phases of the AI system?

Extrait de la liste d'auto-évaluation "The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment" (2020)

La référence aux droits fondamentaux est récurrente dans le cadre de l'évaluation d'impact algorithmique, qu'elle soit spécifiquement traitée ou lapidairement évoquée. Cela conduit à prendre en compte des éléments juridiques, qui revêtent une nature obligatoire et contraignante appelant plutôt à une évaluation de conformité au droit qu'à un exercice d'auto-évaluation fondée sur les risques. Les difficultés d'application pratique des droits fondamentaux dont l'effet direct sur les personnes reste limité, à défaut d'une règlementation plus précise pour leur mise en œuvre, sont souvent citées comme argument pour expliquer la prise en compte de ces droits, en tant que « principes fondamentaux » dans une démarche d'EIA. Le débat reste néanmoins ouvert entre partisans d'une approche basée sur un « principisme éthique » et celle fondée sur le droit[21].

On observe donc une variété des « domaines d'impact » mais reste à comprendre comment les EIA envisagent leur évaluation.

B. Les modalités d'évaluation des impacts : comment est réalisée l'évaluation ?

Les façons d'évaluer les impacts, envisagées par les différentes EIA, semblent varier selon les questions posées qui pourront soit appeler directement à une évaluation au regard des domaines d'impact concernés (i), soit participer à un exercice de déduction des impacts sur la base des caractéristiques des systèmes (ii).

---

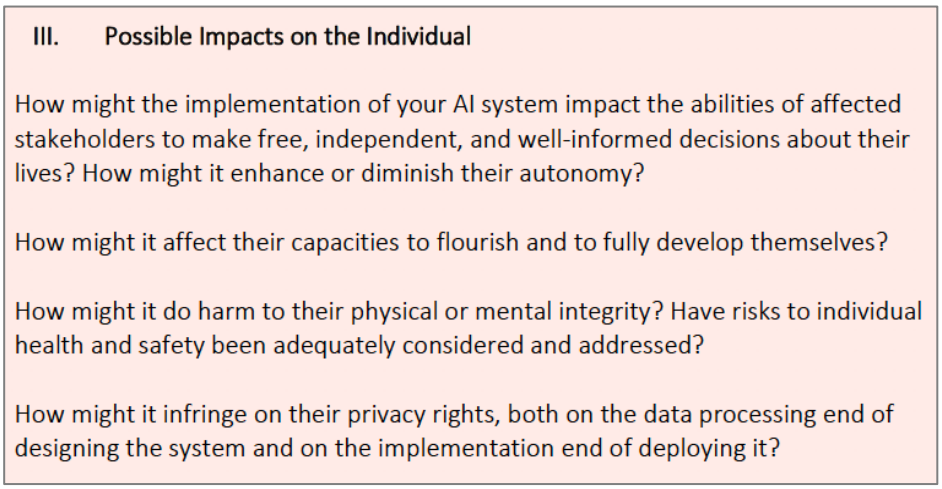
20 L'outil renvoie notamment à la Charte des droits fondamentaux de l'Union européenne et la Convention européenne des droits de l'homme.

21 Wagner, B. (2018). Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping? In M. Hildebrandt (Ed.), Being Profiling. Cogitas ergo sum. Amsterdam University Press.

(i) L'évaluation directe sur la base de questions relatives aux domaines d'impact

Parmi les moyens d'évaluer les impacts, plusieurs EIA intègrent des questions invitant directement les personnes qui réalisent l'EIA à évaluer des impacts sociaux et/ou juridiques.

On retrouve ce type de questions dans plusieurs EIA, comme par exemple, dans le « *Stakeholder Impact Assessment* » ou « *SIA* » proposé au Royaume-Uni, sur la base du rapport produit par The Alan Turing Institute[22], qui invite à s'interroger, entre autres, sur « Comment la mise en œuvre [d'un] système d'IA est-elle susceptible d'impacter les capacités des parties prenantes affectées à prendre des décisions de manière libre, indépendante et éclairée concernant leur vie ? » ou sur la façon dont « les valeurs de participation citoyenne, d'inclusion et de diversité sont prises en compte de manière appropriée dans l'articulation de l'objet et des objectifs du projet [de recours à l'IA] ».



III. Possible Impacts on the Individual

How might the implementation of your AI system impact the abilities of affected stakeholders to make free, independent, and well-informed decisions about their lives? How might it enhance or diminish their autonomy?

How might it affect their capacities to flourish and to fully develop themselves?

How might it do harm to their physical or mental integrity? Have risks to individual health and safety been adequately considered and addressed?

How might it infringe on their privacy rights, both on the data processing end of designing the system and on the implementation end of deploying it?

Extrait du prototype de « Stakeholder Impact Assessment » ou « SIA » (Royaume-Uni)

Les questions peuvent être **ouvertes** (comme celles visées ci-dessus), n'appelant pas de réponses binaires et/ou prédéterminées et invitant plutôt à une évaluation contextuelle et circonstanciée. Des questions plus **fermées** peuvent néanmoins être reprises, notamment dans les EIA plutôt orientées vers une évaluation quantitative de l'impact et du risque. Ces questions fermées vont généralement conduire à des réponses de type « oui/non » ou « cases à cocher ». Par exemple, l'outil canadien va inviter à préciser les « incidences de la décision sur la santé et le bien-être des personnes » en sélectionnant une réponse parmi une liste préétablie de degrés d'impact (de faible à très élevée).

[22] Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector.*The Alan Turing Institute*. https://doi.org/10.5281/zenodo.3240529

De même, l'EIA « *Ethics and Algorithms Toolkit* » va proposer cet exercice sur mode « cases à cocher » pour identifier les impacts des systèmes algorithmiques concernés parmi une liste de domaines d'impact identifiés.


Step 1.1.2 Identify the types of impact
- ❑ Access to goods, benefits or services
- ❑ Financial
- ❑ Property or equipment
- ❑ Reputation
- ❑ Emotional
- ❑ Life / safety
- ❑ Privacy
- ❑ Liberty / freedom
- ❑ Rights / intellectual Property

Extrait de l'outil "Ethics and Algorithms Toolkit" Worksheet for Part 1

Qu'elles soient ouvertes ou fermées, ces questions ont pour point commun de cibler directement un domaine d'impact à évaluer aux fins de l'identifier et/ou de le mesurer. Toutefois, on observe également que certaines EIA vont intégrer des questions de nature différente, semblant participer à une évaluation indirecte des impacts via l'identification de certaines caractéristiques des systèmes algorithmiques.

(ii) L'évaluation indirecte sur la base de questions relatives aux caractéristiques des systèmes

Plusieurs EIA proposent en effet des questions dont l'objet n'est pas l'évaluation directe d'impacts sociaux ou juridiques mais l'identification de caractéristiques associées aux systèmes.

Cette approche est notamment reprise dans l'outil d'auto-évaluation proposé par le groupe d'experts de haut-niveau en IA (GEHN IA) avec des questions telles que « Le système d'IA peut-il affecter l'autonomie humaine en générant une sur-dépendance des utilisateurs ? » ou « Le système d'IA stimule-t-il les interactions sociales avec ou entre les utilisateurs ? ».

Ces questions vont pouvoir porter, par exemple, sur le design du système : « Le système d'IA peut-il générer une confusion pour certains ou tous les utilisateurs sur le fait qu'ils interagissent avec un humain ou une machine ? » (EIA GEHN IA). Elles peuvent également concerner le traitement des données : « Le système exigera-t-il l'analyse de données non structurées pour faire une recommandation ou prendre une décision ? » (EIA Canada) tout comme des aspects organisationnels : « Avez-vous mis en place des mesures pour assurer la traçabilité du système d'IA tout au long de son cycle de vie ? » (EIA GEHN IA) ou « Disposerez-vous de processus documentés pour tester les jeux de données en fonction de biais et d'autres résultats attendus ? » (EIA Canada).

Les réponses à ces questions n'appellent donc pas une évaluation directe des impacts en matière de bien-être ou d'autonomie par exemple. Néanmoins, elles semblent contribuer à cette évaluation de manière indirecte en ce que le risque d'impact sera plus ou moins prononcé selon les réponses données. On peut en effet déduire que dans la mesure où la réponse donnée à la question « Le système d'IA peut-il générer une confusion pour certains ou tous les utilisateurs sur le fait qu'ils interagissent avec un humain ou une machine ? » est « oui », l'impact négatif sur l'autonomie des personnes peut en être déduit.

Cet exercice de déduction semble dès lors nécessiter une expertise préalable permettant d'associer telle ou telle caractéristique liée au système à un risque et/ou impact. Cela implique également que des conclusions ont d'ores et déjà été opérées en amont quant à ces relations de cause à effet. Dès lors, cette expertise est-elle nécessaire pour réaliser une EIA intégrant ce type de questions plus techniques ? Et comment arbitrer la pondération entre ces questions et les questions d'évaluation directe des impacts dans les EIA ?

À ce jour, la variété des contenus des EIA, tant sur l'objet de l'évaluation que sur les effets des systèmes algorithmiques à évaluer, ne permet pas d'identifier une approche « standard ». Là encore, il est difficile d'arbitrer sur ce qui serait le plus adapté et efficace en l'absence d'un retour d'expérience et d'un suivi de leur mise en œuvre pratique. Cependant, une question complémentaire peut se poser concernant l'impact du cadre de ces EIA sur leur efficacité.

C. Le cadre de l'évaluation d'impact algorithmique

Il apparaît en effet que l'EIA est généralement associée à un cadre indiquant son statut obligatoire ou non (i) ainsi que les modalités de réalisation, de communication et de suivi via notamment la publicité de l'EIA (ii).

(i) Caractère obligatoire ou facultatif de l'évaluation d'impact algorithmique

Parmi les différentes EIA examinées, seule l'initiative canadienne s'inscrit dans un **cadre obligatoire règlementaire**. L'outil a en effet été adopté en application de la Directive sur la prise de décision automatisée adoptée par le Secrétariat du Conseil du Trésor du Canada. La Directive, entrée en vigueur le 1er avril 2019, prévoit l'obligation, pour les ministères et services fédéraux, de réaliser une EIA avant le déploiement de tout système décisionnel automatisé développé ou acheté après le 1er avril 2020. Des sanctions peuvent être appliquées en cas de non-conformité à ces exigences.

Cette dimension obligatoire n'est pas évoquée dans les autres exemples d'EIA qui correspondent plutôt à des **démarches internes volontaires**. Cependant, il est intéressant de noter que des projets et recommandations en matière de règlementation future évoquent le caractère obligatoire que pourrait avoir la réalisation d'une EIA. La proposition de loi « *Algorithmic Accountability Act of 2019* » prévoit ainsi la réalisation obligatoire d'une étude d'impact des systèmes de décision automatisée à haut-risque, pour certaines entités de taille et d'importance significative au regard de seuils déterminés. Dans le même esprit, la recommandation du Comité des ministres des États membres invite à rendre obligatoire pour les gouvernements la réalisation d'une étude d'impact sur les droits de l'homme pour tous les systèmes algorithmiques à haut risque dont le déploiement est envisagé. Ces initiatives sont intéressantes à considérer même si l'articulation consistant à conditionner l'application d'une évaluation d'impact à des systèmes nécessitant d'être préalablement évalués (en l'occurrence à **haut-risque**) peut interroger.

Toutefois, au-delà de l'aspect obligatoire ou non de la réalisation d'une EIA, d'autres mesures associées au **cadre organisationnel et procédural** sont régulièrement reprises, à titre de recommandations, pour préciser les modalités liées à la réalisation, la communication, le suivi et le contrôle de l'EIA et de ses résultats. Des recommandations invitant à la **consultation publique** ou la réalisation de l'EIA au stade des **procédures d'achats publics** sont ainsi évoquées mais ce sont les mesures de transparence et de publicité qui reviennent le plus régulièrement et apparaissent revêtir une importance particulière.

(ii) Transparence et publicité des évaluations d'impact algorithmique

L'ensemble des EIA examinées s'inscrivent dans la continuité des principes de transparence et de **redevabilité** consacrés comme « principes éthiques » appliqués au développement et déploiement de systèmes algorithmiques mais également attachés à la transparence de l'action administrative, dès lors que ces systèmes sont utilisés par l'administration.

Des recommandations quasi-unanimes prévues aux EIA associent la réalisation d'une EIA à une publicité des systèmes algorithmiques utilisés et des résultats de l'EIA. L'EIA étant un outil d'auto-évaluation interne, la transparence et la publication sur l'analyse effectuée rend accessible au public des éléments lui permettant de comprendre les systèmes algorithmiques déployés, de les évaluer, de les contrôler et de les contester le cas échéant.

Le « *Stakeholder Impact Assessment* » invite ainsi à partager publiquement les résultats de l'EIA avant le lancement d'un système d'IA pour encourager l'inclusion du public dans l'évaluation des impacts. La recommandation du

Comité des ministres des États membres indique également que les évaluations d'impact sur les droits de l'homme réalisées par les États ou pour leur compte devraient être publiquement accessibles. L'outil proposé par le gouvernement du Canada envisage la publication des « résultats de tout examen ou audit » relatifs aux systèmes décisionnels automatisés revus.

Il apparaît que la publicité de l'EIA peut revêtir des temporalités et des objectifs différents entre acte d'information et acte d'inclusion et de consultation du public. Dans le premier cas, l'EIA, diffusée une fois la décision de déployer le système algorithmique prise, va jouer un rôle d'information du public qui pourra réagir *a posteriori*, le cas échéant. Dans le second cas, l'EIA sera plutôt publiée dans le cadre de la phase de développement du système et avant la décision de le déployer. La publicité sera alors susceptible d'assurer à la fois un rôle d'information et de consultation des personnes affectées et d'inclusion du public.

L'EIA peut donc être considérée comme un moyen de contribuer à la redevabilité sous réserve qu'elle soit réalisée de manière transparente et accessible au public. Cette publicité est d'ailleurs perçue pour beaucoup d'associations de défense des droits sur Internet comme une élément indispensable à l'opposabilité et à l'efficacité réelle d'une EIA. Cette visée de transparence vis-à-vis du public serait-elle ainsi la « raison d'être » de l'évaluation d'impact algorithmique où faut-il percevoir l'EIA comme un outil de sensibilisation interne aux impacts possibles des systèmes algorithmiques sur les personnes et la société ? Au regard des EIA examinées, il paraît difficile d'apporter une réponse tranchée, ces deux objectifs semblant guider le recours à ce type d'outils.

Toutefois, quelle que soit la visée portée par l'EIA, il paraît indispensable d'assurer au préalable une bonne compréhension et accessibilité de l'EIA pour celles et ceux chargées de les réaliser, notamment dans le secteur public. Or, il apparait qu'outre la variété des formes et contenus des EIA, ces outils visent des notions et des concepts dont l'étendue et la complexité amène à s'interroger sur leur « utilisabilité ».

III – L'ÉVALUATION D'IMPACT ALGORITHMIQUE EST-ELLE « UTILISABLE » EN PRATIQUE ?

L'EIA étant un « phénomène récent », la mise en œuvre pratique des outils examinés demeure limitée. Grâce aux différents entretiens réalisés avec plusieurs personnes ayant travaillé à l'élaboration des EIA et/ou suivant leur mise en œuvre, des premiers retours d'expérience et impressions ont pu néanmoins être collectés

et nous verrons qu'au-delà des variétés de formes et de contenus constatés, l'esprit dans lequel ces EIA ont été pensées est souvent similaire. Les informations obtenues ne permettent cependant pas de bénéficier d'une visibilité précise sur l'efficacité et l'utilisabilité des EIA qui ne pourront être appréhendées que via un suivi de leur utilisation à moyen-long terme.

Néanmoins, quelques questions émergent dès à présent, notamment concernant la compréhension même des éléments permettant de réaliser l'évaluation (A). En outre, l'approche d'auto-évaluation des impacts peut amener à s'interroger sur l'efficacité et la pertinence de l'évaluation des systèmes algorithmiques (B).

A. Question de la compréhension : l'évaluation d'impact algorithmique est-elle suffisamment compréhensible pour être utilisable ?

La question de la compréhension peut porter **sur l'objet même des évaluations**. Comme indiqué précédemment, les EIA peuvent avoir pour objet des « systèmes d'IA », des « algorithmes » ou des « systèmes de décision automatisée ». Les notions de systèmes d'IA et d'algorithmes utilisées dans les EIA proposées par le GEHN IA, les gouvernements du Royaume-Uni et de Nouvelle-Zélande n'offrent pas de définitions précises de ces notions. Cela laisse donc une place assez large à l'interprétation tout en nécessitant une certaine expertise en matière d'IA pour en appréhender la technicité. Ce même constat peut s'appliquer aux « systèmes décisionnels automatisés », pourtant définis dans le cadre de l'EIA proposée par le gouvernement du Canada mais dont la définition même en tant que « technologie qui soit informe ou remplace le jugement des décideurs humains »[23] a donné lieu à des questions de la part des utilisateurs de l'EIA sur l'identification des systèmes algorithmiques pouvant entrer (ou non) dans ces catégories.

Les difficultés de compréhension de l'objet de l'évaluation paraissent ainsi susceptibles de provoquer un premier frein à l'accessibilité de l'EIA. C'est un point d'attention également identifié par les services des gouvernements ayant proposé des EIA, sur la base des premiers retours reçus.

Cette question de la compréhension peut également porter **sur les domaines d'impact**. Comme indiqué précédemment, les effets des systèmes algorithmiques à évaluer peuvent concerner différents types d'impacts, comme par exemple l'impact sur le bien-être, la vie privée ou les biais et discriminations. L'évaluation de ces impacts suppose de comprendre à quoi ils correspondent au regard des domaines visés. Ainsi, lorsqu' il est demandé d'évaluer l'incidence d'un système

[23] Directive sur la prise de décision automatisée – Annexe A - Définitions

de décision automatisé « sur la santé et le bien-être », comme cela est le cas, par exemple, dans l'EIA du Canada, il est probable que les personnes réalisant l'EIA seront amenés à s'interroger sur ce qu'est le bien-être et comment le mesurer, au regard de quels critères.

La difficulté d'interprétation est accrue d'autant que ces notions relèvent de conceptions culturelles et locales distinctes les unes des autres. Ces disparités peuvent ainsi questionner l'application uniforme d'EIA au niveau international sauf à envisager l'éventualité et la faisabilité d'une « standardisation » sur la base d'un consensus d'indicateurs communs.

Il apparaît donc qu'en l'absence de définitions et indications claires et face à l'ampleur et la complexité des domaines d'évaluation à considérer, l'utilisabilité de l'EIA risque d'être limitée. La Commission européenne, dans son Livre Blanc sur l'IA, indique d'ailleurs que les « éléments permettant d'établir qu'une application d'IA est à haut risque devraient être clairs, faciles à comprendre et applicables à toutes les parties concernées ».

Lors des entretiens menés avec différents interlocuteurs des services gouvernementaux du Royaume-Uni, du Canada et de Nouvelle-Zélande, la difficulté de la compréhension de certaines notions a été remontée suite aux initiatives de mise en œuvre initiées par quelques départements et services et certaines mesures sont d'ores et déjà considérées pour tenter d'y remédier comme notamment :

- l'**élaboration de documents de travail** destinés à apporter des définitions et informations complémentaires pour répondre aux questions les plus fréquentes et fournir des précisions sur la logique et méthode d'évaluation. Le Conseil de Trésor du Canada travaille par exemple sur une nouvelle page d'accueil de l'EIA dont la publication est prévue prochainement et incluant une explication du système de pointage de l'EIA ;
- la tenue d'**ateliers pluridisciplinaires** dès la conception de l'EIA puis dans le cadre de sa mise œuvre, pour tester la compréhension et l'utilisabilité des outils proposées et les adapter en conséquence ;
- la constitution d'**équipes pluridisciplinaires** pour la réalisation de l'EIA et permettant de réunir les différentes expertises nécessaires à l'appréhension des différents domaines d'impact (par exemple juridique, technique, sociologique, etc.) ; ou

- la **formation** des agents publics. L'axe de formation des agents publics est d'ailleurs un objectif expressément visé dans la Directive sur la prise de décision automatisée adoptée au Canada[24].

La compréhension des éléments permettant la réalisation de l'EIA représente ainsi un enjeu essentiel pour l'accessibilité et l'utilisabilité de ces outils. Les difficultés pratiques sont réelles mais il est néanmoins intéressant de mentionner la dimension plus large de l'EIA comme outil de **sensibilisation** si ce n'est à l'identification précise des effets des systèmes algorithmiques mais tout au moins aux effets qu'ils peuvent avoir. La grande majorité des personnes interrogées lors de cette étude a insisté sur cet aspect et sur le fait que l'EIA, loin d'être un outil auto-suffisant, participe avant tout à un travail d'**acculturation** destiné à accompagner progressivement l'intégration des questions sur les effets sociétaux des systèmes algorithmiques, dans leur pratique. L'EIA est un moyen de créer le dialogue avec les équipes des différentes administrations voire de générer une communauté mettant en lien différents « utilisateurs » de l'EIA pour soutenir la dynamique d'appropriation de l'outil comme cela a été souligné par les membres de l'équipe de Stats NZ.

L'EIA viserait ainsi, à plus ou moins long terme, une montée en compétences et en autonomie des agents publics concernant l'identification des effets des systèmes algorithmiques. Un accompagnement initial des équipes paraît néanmoins nécessaire tout en prenant compte des limites de moyens et de temps pouvant être dédiés à cet effet.

La compréhension de ce qui permet d'évaluer les systèmes algorithmiques est donc clé pour son utilisabilité, d'autant plus lorsque l'évaluation se base essentiellement sur un auto-arbitrage des effets à considérer.

B. Question de l'auto-arbitrage : l'autonomie dans l'évaluation de l'impact algorithmique nuit-elle à son utilisabilité ?

Comme nous l'avons vu précédemment, les EIA regorgent de notions complexes, plus ou moins définies, laissant la place à une **marge d'interprétation sur l'évaluation même des impacts**. Par exemple, une question telle que « Le système d'IA peut-il avoir un impact négatif sur la société dans son ensemble ou la démocratie ? » laisse une large part, sans autre précisions, à l'auto-arbitrage pour la détermination de l'impact selon sa propre interprétation de concepts qui demeurent flous et difficilement accessibles dans leur compréhension.

[24] L'article 6.3.5 de la Directive prévoit l'obligation de « Fournir aux employés la formation appropriée concernant la conception, la fonction et la mise en œuvre du système décisionnel automatisé afin d'être en mesure d'examiner, d'expliquer et de surveiller le fonctionnement du système, conformément à ce qui est prévu à l'annexe C. »

L'évaluation porte également le plus souvent sur des éléments non-mesurables (ou tout au moins non-calculables) sans qu'une seule réponse précise puisse être apportée.

Dès lors, moins les domaines d'impact seront définis ou définissables, plus ils seront sujets à l'interprétation non guidée des personnes chargées de les évaluer laissant la place à des visions personnelles et appréhensions du risque qui le seront tout autant. Un manque de cohérence dans l'application de l'EIA peut donc être à craindre dans les évaluations réalisées. À cela s'ajoute la question de l'**exercice d'auto-évaluation interne** où les personnes réalisant l'EIA sont généralement à la fois juges et parties.

Afin de limiter ces effets susceptibles d'affaiblir la crédibilité et la pertinence de l'EIA, plusieurs recommandations ont été émises pour renforcer la redevabilité, notamment des administrations. Parmi ces recommandations figurent :

- la transparence et la publicité de l'EIA avant et/ou après la décision de déploiement du système algorithmique ;
- l'inclusion des parties prenantes pendant la réalisation de l'EIA pour l'identification et l'évaluation des impacts des systèmes algorithmiques pouvant les affecter ; ou
- la revue et/ou le contrôle des systèmes (et de l'EIA) par des pairs en interne (comme c'est le cas pour l'EIA proposée par le gouvernement du Canada) ou des experts externes (comme recommandé par l'AI Now Institute dans son rapport).

Ces recommandations renvoient à celles de transparence et de publicité qui font de l'EIA un outil tant destiné à l'usage interne des agents publics qu'à l'implication du public et/ou d'experts externes. C'est aussi ce qui permettrait de rééquilibrer le pouvoir d'appréciation laissé pour opérer les évaluations, grâce à un suivi et contrôle extérieur. L'avenir nous dira si cette approche transparente est retenue et l'impact que cela pourra avoir pour favoriser (ou non) la réalisation d'EIA même s'il apparaît d'ores et déjà difficile d'assurer la crédibilité de l'EIA vis-à-vis du public sans transparence et inclusion des parties prenantes pour la réalisation de l'EIA.

CONCLUSION

Les évaluations d'impact algorithmique en sont donc à leurs débuts. Le recul manque encore pour en appréhender l'intérêt et l'efficacité et un suivi de leur mise en œuvre paraît indispensable. Difficile également à l'heure actuelle de préjuger de l'utilisabilité et de la pertinence d'une approche d'EIA par rapport à une autre même si l'on perçoit que l'inclusion des parties prenantes et du public dans le processus d'évaluation via la publicité de l'EIA semble critique pour « valider » l'auto-évaluation d'impact.

Mais avant même la réalisation de cet idéal de transparence, la compréhension des éléments permettant de réaliser l'EIA paraît être un préalable nécessaire à toute mise œuvre effective. Comme nous l'avons vu, cette compréhension n'est pas aisée à assurer, notamment au regard de la diversité des outils et des notions parfois floues et complexes qu'ils contiennent et qui ne permettent pas une lisibilité précise quant à ce qu'est une EIA et comment la réaliser.

Dès lors, faudrait-il envisager l'EIA sous l'angle d'une initiative élargie, possiblement internationale, tournée vers un travail de « standardisation » de l'EIA afin d'en acter une compréhension commune ? Cette démarche est-elle cependant réaliste dès lors que des concepts largement interprétables sont concernés ? Ces questions amènent également à s'interroger sur l'opportunité d'une approche plus sectorielle ou ciblée des EIA en tant qu'outils « sur-mesure », adaptés à des contextes d'utilisation particuliers pouvant influencer l'identification des impacts et de leur portée. Deux approches se distinguent en effet entre l'évaluation uniforme d'un algorithme, quel que soit son cadre de mise en œuvre et une évaluation différenciée d'un même algorithme selon le secteur de déploiement. Il sera intéressant d'observer l'application des EIA par les différentes administrations afin d'observer dans quelle mesure ces dernières s'approprieront ces outils en adaptant, le cas échéant, leur contenu par rapport à leurs activités.

Il apparaît donc difficile de dégager une conclusion précise quant aux modalités de mises en œuvre des EIA. Cela vaut également concernant leur statut. Nous avons en effet vu que les EIA sont envisagées comme des outils participants à une acculturation aux effets sociétaux des systèmes algorithmiques. Doit-on s'en tenir à cet objectif ou envisager également une portée plus structurante en intégrant les EIA à une règlementation plus précise pour encadrer voire arbitrer le recours à des systèmes algorithmiques ? Il pourrait être dès lors intéressant de préciser le statut de l'EIA au regard de la décision même de déployer ou non un système algorithmique au regard des impacts identifiés. En effet, cette question de l'opportunité du recours à un algorithme n'apparait pas précisément liée à l'exercice d'évaluation d'impact algorithmique qui reste

souvent perçu comme un outil d'évaluation de risques dans le but d'identifier des mesures de limitation de ces risques pour permettre le déploiement du système. Mais qu'en est-il lorsque les risques identifiés sont tels qu'ils apparaissent difficiles si ce n'est impossibles à limiter ? L'EIA pourrait-elle conditionner le recours à un système algorithmique comme envisagé par le Conseil des ministres aux États membres dans sa recommandation qui propose que, dès lors, qu'une étude d'impact sur les droits de l'homme « permet d'identifier des risques importants qui ne sauraient être atténués, le système algorithmique ne devrait être mis en œuvre ou utilisé par aucune autorité publique » ?

Autant de questions qui s'attachent à appréhender l'utilité et l'efficacité réelle des EIA et qui restent à explorer. Leur portée et crédibilité en tant qu'outil participant à la transparence et à la redevabilité, notamment dans le secteur public, en dépendra. Cependant, à ce jour, les EIA proposées ont le mérite d'exister et si l'évaluation d'impact algorithmique doit encore faire ses preuves, elle contribue à une prise de conscience nécessaire des enjeux sociétaux des systèmes algorithmiques.

REMERCIEMENTS

Cette étude s'est indéniablement enrichie des différents entretiens réalisés avec des interlocutrices et interlocuteurs ayant participé à l'élaboration et/ou suivant l'utilisation d'outils d'évaluation d'impact algorithmique. Un grand merci à Natalia Domagala, Benoît Deshaies, David Leslie, Hubert Guillaud, Mike Katell, Amba Kak, Jeanne McKnight, Alessandro Aduso, Matthias Spielkamp, Fanny Hidvégi, Daniel Leufer et Helen Darbishire pour avoir partagé leurs expériences et réflexions sur ce sujet. Et bien évidemment, merci à Soizic Pénicaud et Simon Chignard du département Etalab pour leur accompagnement.