

INTELLIGENCE ARTIFICIELLE ET TRANSFORMATION DE L'ÉVALUATION DE PROGRAMME

Steve Jacob

Seima Souissi

Loïc Duplantis



UNIVERSITÉ
LAVAL

Chaire de recherche sur l'administration
publique à l'ère numérique

En partenariat avec :



Steve Jacob, Seima Souissi et Loïc Duplantis
Intelligence artificielle et transformation de l'évaluation de programme
Chaire de recherche sur l'administration publique à l'ère numérique
Université Laval
Québec, 2023

Les analyses et commentaires présentés dans ce document n'engagent que leurs auteurs et ne constituent pas une position officielle.

Les auteurs

Steve Jacob est professeur titulaire de science politique à l'Université Laval, où il dirige le laboratoire de recherche sur la performance et l'évaluation de l'action publique (PerfEval). Il est titulaire de la Chaire de recherche sur l'administration publique à l'ère numérique et il codirige la fonction Politiques publiques de l'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique, une division visant à apporter un soutien scientifique et analytique aux décideurs publics. Steve Jacob mène des recherches sur les processus de modernisation de l'administration, l'éthique publique et les dispositifs d'évaluation et de gestion de la performance.

Seima Souissi est titulaire d'un doctorat en communication publique de l'Université Laval. Elle est professionnelle de recherche à la Chaire de recherche sur l'administration publique à l'ère numérique et chargée de cours au Département d'information et de communication de l'Université Laval.

Loïc Duplantis est étudiant au baccalauréat intégré en économie et politique à l'Université Laval. Il est auxiliaire de recherche à la Chaire de recherche sur l'administration publique à l'ère numérique.

Chaire de recherche sur l'administration publique à l'ère numérique

En partenariat avec le **Secrétariat du Conseil du trésor**, la Chaire de recherche sur l'administration publique à l'ère numérique a pour mission de produire des connaissances de pointe sur les effets des transformations numériques et sur les défis que pose cette transformation à l'administration publique. La Chaire est propulsée par l'**Académie de la transformation numérique (ATN)**. En plaçant l'humain au cœur de sa démarche, l'ATN permet aux employés et aux gestionnaires d'acquérir les connaissances et de développer les compétences nécessaires pour relever les défis que soulève l'arrivée massive du numérique. Elle offre aux organisations d'assumer un véritable rôle de bâtisseur en s'impliquant activement dans la cocréation de programmes de formation multidisciplinaires ancrés dans leur nouvelle réalité numérique.

www.administration-numerique.chaire.ulaval.ca



Table des matières

Introduction	2
1. Usages possibles de l'IA en évaluation	3
1.1. L'IA pour préparer l'évaluation	3
1.1.1. Délimitation du périmètre de l'évaluation.....	3
1.1.2. Sélection des documents	4
1.1.3. Analyse des sentiments.....	4
1.2. L'IA pour l'exploration des données d'évaluation.....	5
1.3. L'IA pour l'étude des inférences causales	6
1.4. L'IA pour l'analyse avancée de contenu.....	6
1.5. L'IA pour évaluer les effets des programmes.....	7
2. L'IA pour moderniser le cycle des politiques publiques	8
3. Les défis de l'IA en évaluation de programme	10
3.1. Les défis de l'évaluation continue des politiques publiques.....	10
3.2. L'usage responsable et éthique des données	10
3.3. La transformation du métier et des compétences.....	11
Conclusion	12
Bibliographie	13

Introduction

L'évaluation de programmes et de politiques publiques vise à évaluer les conséquences prévues et imprévues des interventions publiques pour savoir si elles ont atteint leurs objectifs (Jasper et coll., 2019). Le Secrétariat du Conseil de trésor (SCT) du Québec définit l'évaluation comme une démarche systématique de collecte et d'analyse de données et de renseignements sur les programmes en vue d'améliorer ceux-ci ou de poser un jugement sur leur valeur et, ainsi, d'éclairer la prise de décision. L'évaluation comprend notamment les travaux portant sur les besoins ou les moyens d'intervention, ou encore sur l'implantation ou la mise en œuvre des programmes (SCT, s.d.).

Les activités de recherche et d'analyse se trouvent au cœur du processus d'évaluation. Elles s'appuient sur l'examen de vastes ensembles d'informations contenues généralement dans les rapports et les documents produits dans le cadre des programmes (Jasper et coll., 2019). De nombreux auteurs laissent alors entendre que puisque l'intelligence artificielle (IA) et l'analyse de données massives ont déjà transformé les domaines de la recherche dans de multiples disciplines, elles posséderaient un potentiel important pour améliorer l'analyse des politiques publiques et transformer la pratique de l'évaluation au sein des organisations publiques (Pencheva et coll., 2020).

L'application de différentes technologies de l'IA, comme l'apprentissage automatique, le traitement du langage naturel ou la vision par ordinateur, à la collecte et à l'analyse des données est susceptible d'accélérer le processus d'évaluation et d'offrir de nouvelles perspectives à la pratique évaluative.

Néanmoins, la littérature révèle que, pour l'instant, l'usage de l'IA dans l'évaluation de programme se situe encore au stade de l'exploration. Des travaux pionniers tentent de déterminer l'utilité et l'efficacité d'introduire ces technologies aux fins d'évaluation. Considérant la capacité de ces outils à analyser de grands volumes de données de manière quasi instantanée, des auteurs estiment que l'IA révolutionnera la conception traditionnelle de l'évaluation en tant qu'étape qui marque la fin du cycle de la politique publique. La notion d'évaluation fera désormais partie intégrante de toutes les étapes du cycle des politiques publiques. Le traitement des données massives et l'analyse prédictive fondée sur l'IA accompagneront les responsables politiques dès la phase de l'élaboration des politiques. De cette façon, ces derniers pourront prendre des décisions éclairées et ajuster leurs interventions en fonction des données recueillies, au fur et à mesure de la mise en œuvre des politiques.

Dans le présent rapport, nous passons d'abord en revue les principaux usages possibles de l'IA en matière d'évaluation, identifiés par les études exploratrices. Ensuite, nous mettons l'accent sur le potentiel de l'IA d'introduire l'évaluation continue dans le cycle des politiques publiques ainsi que les avantages et les inconvénients que cette nouvelle approche pourrait entraîner.

1. Usages possibles de l'IA en évaluation

À défaut de trouver des publications qui portent sur les usages effectifs des technologies de l'IA par les organisations publiques aux fins d'évaluation de programme, nous avons identifié des études qui explorent les applications possibles de ces technologies dans ce domaine. Ces études sont axées davantage sur les explications techniques quant au développement des solutions et des algorithmes d'IA. Cependant, les fruits de ce travail de recherche et de développement sont ensuite testés, pour la plupart des études, sur des projets d'évaluation de programme concrets.

Dans les paragraphes suivants, nous présentons ces applications possibles de l'IA dans le processus d'évaluation de programme ainsi que dans le contexte des projets d'évaluation qui ont permis de vérifier leur applicabilité et leur efficacité. Nous exposons aussi les principaux résultats tirés de ces études exploratoires.

1.1. L'IA pour préparer l'évaluation

Dans une étude commanditée par la Direction de l'évaluation de l'aide internationale (PRA) d'Affaires mondiales Canada (AMC), Jasper et coll. (2019) analysent les possibilités de recourir aux approches modernes de science des données dans le processus d'évaluation de programme.

Partant d'une revue de littérature, les auteurs explorent l'utilisation des méthodes d'analyse de texte soutenues par l'IA ainsi que les techniques d'apprentissage automatique et de traitement du langage naturel (*Natural Language Processing – NLP*), afin d'améliorer le processus avec lequel AMC met en œuvre les évaluations de programme. L'objectif global de ce travail d'exploration est de fournir une description des usages possibles de ces méthodes, ainsi que les avantages et les défis que ces usages pourraient représenter dans le contexte d'évaluation de programme.

Après avoir examiné le processus actuel par lequel AMC met en œuvre les évaluations de programme, les auteurs déterminent trois étapes principales du processus d'évaluation dans lesquelles les méthodes modernes de la science des données pourraient être appliquées de manière utile et efficace. Ces étapes sont : 1) la délimitation du périmètre de l'évaluation, 2) la sélection des documents à inclure dans l'examen des projets identifiés et 3) le codage et l'analyse des documents. L'étude montre qu'une gamme d'approches différentes de NLP pourraient être appliquées à ces trois étapes d'évaluation de programme (Jasper et coll., 2019).

1.1.1. Délimitation du périmètre de l'évaluation

Les programmes d'AMC financent un grand nombre de projets qui ne peuvent pas tous être inclus dans une évaluation. La sélection des projets s'avère donc une étape importante dans toute évaluation de programme. AMC sélectionne ces projets en partie en fonction des domaines thématiques d'intervention dont ils relèvent. Dans la pratique actuelle, les évaluateurs lisent la documentation du projet et extraient manuellement l'information nécessaire pour déterminer l'orientation thématique d'un projet. Selon une estimation fournie par AMC, pour un évaluateur, l'examen de tous les documents de projet liés aux plus récents programmes d'AMC en matière de santé des mères, des nouveau-nés et des enfants nécessiterait plus de 1 000 ans.

L'étude montre qu'une grande partie de ce processus peut être automatisée et améliorée grâce à la modélisation thématique. Cette méthode basée sur le traitement du langage naturel pourrait servir à identifier automatiquement les thèmes ou les sujets non observés dans de grands ensembles de données textuelles et, ensuite, à étiqueter les documents en conséquence (Jasper et coll., 2019).

1.1.2. Sélection des documents

Dans le processus actuel d'évaluation des programmes d'AMC, les évaluateurs sélectionnent un ensemble prédéfini de documents à examiner. Or, selon cette étude, plusieurs méthodes de récupération d'informations peuvent être utilisées pour repérer automatiquement les renseignements recherchés (p. ex. des noms ou des mots) dans de grands ensembles de données textuelles. Les systèmes de requête qui autorisent les opérateurs logiques dans les bases de données textuelles en constituent un exemple. Un autre exemple est « Word2Vec », une méthode qui utilise l'apprentissage automatique pour estimer la signification des mots en fonction de la façon dont ils coexistent avec d'autres mots. Enfin, la reconnaissance d'entité nommée (NER) est une approche basée sur l'apprentissage automatique qui peut être utilisée pour étiqueter rapidement de gros volumes de texte selon la présence ou non de ces entités (p. ex. des noms d'organisations ou d'individus). Selon cette étude, ces techniques de recherche d'information peuvent contribuer à améliorer la rapidité et la facilité avec lesquelles les documents sont sélectionnés dans les évaluations de programme, et à éviter que les documents contenant des informations cruciales ne soient négligés (Jasper et coll., 2019).

1.1.3. Analyse des sentiments

Les techniques avancées d'IA, telles que le traitement du langage naturel, les techniques d'apprentissage automatique et l'analyse de texte, sont utilisées également pour identifier les émotions exprimées dans un document. Ces méthodes d'analyse sont basées sur des dictionnaires prédéfinis de mots associés à des émotions positives, neutres ou négatives et elles peuvent être appliquées à de grandes quantités de texte. Le recours à ces méthodes en contexte d'évaluation permettrait de discerner les sentiments exprimés à l'égard d'un projet particulier, ce qui aiderait les évaluateurs à déterminer les défis de mise en œuvre, l'impact et la pertinence des programmes (Jasper et coll., 2019).

Pour illustrer ce que l'usage de ces méthodes pourrait donner dans la pratique, Jasper et coll. (2019) présentent une étude de cas qui montre comment la recherche d'information, la modélisation thématique et l'analyse des sentiments ont été appliquées dans un projet visant à analyser les opinions publiques en Ouganda. L'étude de cas porte sur les premiers débats présidentiels télévisés transmis en direct en prélude aux élections générales qui ont eu lieu en 2016. Elle avait pour objet d'identifier les perceptions du public sur la façon dont les débats ont été organisés et si ceux-ci sont considérés comme pertinents pour le processus électoral. L'étude s'est penchée sur les médias sociaux (Facebook), où les débats ont été largement discutés. La réalisation de cette étude impliquait la collecte, la compilation et l'analyse d'un grand volume de textes non structurés. L'étude a utilisé différentes méthodes de NLP pour analyser les données textuelles provenant des commentaires des citoyens. Ces commentaires ont d'abord été filtrés à l'aide de systèmes de recherche d'information par mots-clés. Ensuite, la modélisation thématique a été utilisée pour déterminer les sujets clés et catégoriser les commentaires par rapport aux sujets qu'ils traitaient. Deux grandes catégories ont émergé : « commentaires généraux » et « commentaires sur des

sujets spécifiques ». Pour cette deuxième catégorie, les principaux sujets identifiés concernaient les « candidats », les « modérateurs », les « organisateurs » et la « sensibilisation ».

Dans une dernière étape, l'analyse des sentiments a été utilisée pour évaluer les tendances à l'égard de ces sujets. L'étude a révélé, entre autres, que plus de 90 % des discussions autour des modérateurs étaient négatives. De plus, le public s'est dit préoccupé par le choix des modérateurs, la conduite de certains modérateurs et la façon dont ils ont posé des questions.

L'étude a ainsi permis d'illustrer comment une combinaison de méthodes de récupération d'informations et des techniques avancées d'apprentissage automatique peuvent être utilisées pour améliorer l'efficacité et l'efficacité de la catégorisation et de l'analyse de gros volumes de texte en langage naturel non structuré (Jasper et coll., 2019).

1.2. L'IA pour l'exploration des données d'évaluation

Parmi les projets exploratoires que nous avons repérés, l'étude de cas réalisée par Beitel (2005) examine l'exploitabilité des outils d'exploration de données par l'IA à des fins d'évaluation de programme en milieu scolaire.

L'étude se propose de tester l'applicabilité de ce type de logiciels pour analyser les données relatives au milieu scolaire (les résultats de tests standardisés, les évaluations effectuées par les enseignants, les informations démographiques, etc.). Elle vise également à éclairer les responsables du domaine de l'éducation sur la faisabilité et l'utilité de ces méthodes pour évaluer les programmes et améliorer le rendement des élèves (Beitel, 2005).

L'analyse quantitative d'exploration de données a été réalisée à l'aide du logiciel SPSS pour extraire les éléments de données requis qui étaient enregistrés dans Excel. Puis, ces éléments ont été exportés à partir d'Excel pour être analysés au moyen de la fonction *SPSS Tree* et de DTREG, un outil d'exploration de données d'IA. L'analyse prédictive de DTREG a été menée pour connaître le meilleur prédicteur de la réussite en lecture. À cette fin, trois types de variables disponibles dans la base de données ont été utilisés. La puissance analytique de DTREG a permis de déterminer les prédicteurs de la réussite des élèves dans les tests de lecture (Beitel, 2005).

Le projet a fourni des analyses significatives et stimulantes aux responsables de l'éducation. L'utilisation d'outils d'exploration de données d'IA permet d'approfondir le niveau d'analyse en offrant au personnel de l'éducation la possibilité de répondre à des questions complexes concernant l'impact des programmes et de leur efficacité. L'étude conclut que les décideurs publics peuvent soutenir leurs décisions visant à éliminer, poursuivre ou étendre des programmes particuliers dans leurs établissements avec l'utilisation de ces outils d'IA (Beitel, 2005).

1.3. L'IA pour l'étude des inférences causales

Parmi les travaux qui explorent le potentiel de l'IA dans le domaine de l'évaluation des politiques, plusieurs considèrent les possibilités de recourir aux méthodes d'apprentissage automatique pour l'étude des inférences causales, c'est-à-dire le processus par lequel on peut établir une relation de causalité entre un élément et ses effets (Athey et Imbens, 2017; Kreif et DiazOrdaz, 2019). L'analyse de l'inférence causale, qui relève des techniques d'IA, offre une compréhension plus profonde que la simple étude des corrélations à laquelle procède classiquement la science statistique (Alcaud, 2021).

La plupart des questions posées en évaluation des politiques sont de nature causale, même si elles ne sont pas spécifiquement formulées ainsi. C'est pourquoi les chercheurs qui pratiquent l'évaluation sont préoccupés par des questions contrefactuelles, à savoir ce qui se serait passé en l'absence d'une politique. Cependant, le problème fondamental de l'inférence causale est le suivant : les faits contrefactuels ne peuvent jamais être observés directement. Cette situation rend difficile l'application à l'inférence causale les méthodes d'apprentissage automatique principalement destinées à la prédiction (Kreif et DiazOrdaz, 2019).

Récemment, de nouvelles études ont commencé à examiner l'utilité de l'apprentissage automatique pour les problématiques causales, comme l'étude de Kreif et DiazOrdaz (2019). Les auteurs y présentent un exemple de l'usage des méthodes d'apprentissage automatique pour répondre à des questions causales typiques, dans le contexte d'évaluation de programme. L'étude vise à montrer comment l'inférence causale peut être mise en évidence avec des logiciels statistiques largement utilisés. Les auteurs illustrent leur méthode à travers une étude de cas d'évaluation d'un régime d'assurance maladie sur l'utilisation des soins de santé en Indonésie. L'étude de cas présente des caractéristiques typiques des évaluations appliquées : un résultat binaire et un mélange de covariables binaires, catégorielles et continues (Kreif et DiazOrdaz, 2019).

1.4. L'IA pour l'analyse avancée de contenu

En plus des travaux qui se sont concentrés sur l'application de l'IA pour la collecte, le nettoyage et la modélisation des données à des fins d'évaluation, d'autres études ont exploré l'utilisation de ces technologies pour la synthèse de contenu et les analyses théoriques en contexte d'évaluation (Frazen et coll., 2022).

Le volume et l'accessibilité croissants des données d'évaluation de projet représentent à la fois une solution et un défi pour offrir une analyse approfondie sur la performance des programmes. Bien que les méthodes d'évaluation actuelles puissent s'avérer très efficaces pour générer des données pertinentes sur les programmes, la nature manuelle des analyses traditionnelles empêche l'examen complet et systématique de volumes importants de données d'évaluation (Frazen et coll., 2022).

Les praticiens de l'évaluation ne peuvent pas synthétiser efficacement des bases de données massives au moyen des outils traditionnels d'évaluation. Ils ne peuvent profiter de l'ampleur et de la richesse que les données massives sont susceptibles d'offrir. Puisque l'application des technologies de l'IA au domaine de l'évaluation a révolutionné la capacité à **donner un sens aux données massives, elle pourrait offrir une synthèse beaucoup plus rapide, plus complète et systématique des données et enrichir les résultats des**

évaluations. De plus, ces analyses automatisées pourraient être régulièrement mises à jour, à mesure que de nouvelles données et connaissances sont générées, ce qui garantirait l'actualisation continue des connaissances sur le développement des projets (Frazen et coll., 2022).

L'article de Frazen et coll. (2022) présente la méthodologie et les résultats d'un projet pilote qui avait pour objet de tester l'applicabilité, l'utilité et la valeur ajoutée de l'utilisation de l'IA pour l'analyse avancée du contenu basée sur la théorie. Commandité par le Groupe indépendant d'évaluation (GEI) de La Banque mondiale, ce projet visait à vérifier si les technologies de l'IA pouvaient soutenir et accélérer l'évaluation du rendement des projets et des activités de La Banque mondiale réalisés sur des périodes de 5 à 10 ans (Frazen et coll., 2022).

Traditionnellement, la synthèse qualitative était utilisée pour effectuer une analyse structurée des rapports de projets axés sur la théorie. Ce projet pilote visait à évaluer les gains d'efficacité générés par **l'analyse de contenu assistée par l'IA dans l'étiquetage et la classification du texte, selon un cadre conceptuel axé sur les résultats**. Les chercheurs voulaient vérifier si des modèles de classification de textes peuvent être utilisés pour reproduire avec précision l'analyse manuelle de contenu, et déterminer si ces modèles peuvent effectuer des analyses rapides et efficaces. Pour ce faire, ils ont appuyé l'étude sur les rapports de projets d'un ensemble d'interventions de La Banque mondiale pour la lutte contre le retard de croissance et la malnutrition chronique. Le corpus comprenait 392 rapports de projets uniques émanant de 64 pays.

L'expérience a combiné les techniques d'apprentissage automatique supervisé et non supervisé. Dans l'ensemble, le modèle thématique développé a montré d'excellentes performances dans l'identification des sujets inductifs qui non seulement étaient nouveaux et pertinents pour le domaine, mais se sont révélés des prédicteurs clés de la performance des projets de développement et des bonnes pratiques. La méthode de visualisation utilisée s'est aussi avérée efficace pour identifier des tendances importantes dans les données, tendances qu'un analyste humain n'aurait pu déceler aisément, ce qui a facilité l'établissement de caractéristiques uniques des programmes par pays. Des travaux supplémentaires seraient nécessaires pour optimiser le modèle et renforcer sa performance (Frazen et coll., 2022).

Les auteurs concluent que les résultats de cette étude sont prometteurs. Ceux-ci contribuent à un corpus de travaux sur l'utilisation de l'IA à des fins de synthèse d'évaluation qui se trouvent encore à un stade précoce.

1.5. L'IA pour évaluer les effets des programmes

Dans un article portant sur la possibilité d'utiliser les technologies de l'IA pour évaluer les programmes de santé numériques en Inde, Mohan et coll. (2019) étudient la possibilité d'appliquer les algorithmes d'apprentissage automatique pour évaluer l'efficacité de deux initiatives de santé mobile implantées dans de nombreux États indiens depuis leur lancement en 2012, et pour améliorer leur performance. La première initiative consiste en un service de messagerie audio nommé *Kilkari* destiné à encourager les femmes enceintes à assister aux activités des centres de santé primaire de proximité, où des infirmières et des sages-

femmes recueillent de l'information sur l'évolution de leur grossesse. Les messages hebdomadaires sont adaptés à l'âge gestationnel de l'embryon. Le deuxième programme, *Mobile Academy*, est un cours de formation audio interactif offert par téléphone aux agents de santé de première ligne. Il vise à rafraîchir leurs connaissances sur les comportements de santé préventifs et à améliorer leurs compétences en communication interpersonnelle (Mohan et coll., 2019).

Les auteurs décrivent et expliquent le processus de développement de l'outil d'IA basé sur l'apprentissage automatique qui va rendre possible cette évaluation de programme. Ils traitent de la constitution des bases de données jusqu'aux tests de validation, en passant par les étapes d'élaboration des algorithmes.

Les deux études de cas montrent que la mise en œuvre appropriée de méthodologies d'apprentissage automatique peut générer des données probantes sur l'efficacité du programme et faciliter des interventions dynamiques en temps réel pour améliorer leur mise en œuvre. Le système d'évaluation développé vise à aller au-delà de la compréhension de la performance technologique des programmes. Il cherche à évaluer l'engagement des utilisateurs ainsi que les principales mesures de performance du programme, comme l'achèvement de la formation pour *Mobile Academy* et la couverture de l'exposition des utilisateurs pour *Kilkari* (Mohan et coll., 2019).

2. L'IA pour moderniser le cycle des politiques publiques

Selon l'approche séquentielle des politiques publiques, l'évaluation de programme arrive à la toute fin du cycle des politiques publiques constitué de plusieurs étapes : 1) la mise à l'ordre du jour, 2) la formulation des politiques et la prise de décisions, 3) la mise en œuvre ou l'implantation des politiques et 4) l'évaluation des politiques. Cependant, une politique publique n'est pas achevée une fois que ses résultats sont analysés et évalués. Cette analyse peut conduire à une nouvelle définition du problème et provoquer le redémarrage du cycle (Valle-Cruz et coll., 2020). Toutefois, la boucle de rétroaction dans le processus d'évaluation traditionnel est lente et les résultats prennent beaucoup de temps avant d'être disponibles. Certains auteurs suggèrent alors que le recours aux données massives et aux technologies de l'IA transformera la notion d'évaluation grâce au traitement quasi instantané des données issues de sources multiples. Les résultats d'évaluation peuvent être transmis au fur et à mesure que les données sont collectées (Höchtel et coll., 2016; Pencheva et coll., 2020).

Selon les auteurs, avec de telles capacités, l'IA et les mégadonnées faciliteront l'intégration de l'évaluation dans toutes les phases du cycle des politiques publiques. Elles remplaceront ainsi l'évaluation de fin de cycle par la notion d'évaluation continue des programmes publics. Au lieu d'être une étape bien définie qui se produit à la suite de l'étape de la mise en œuvre, l'évaluation sera désormais intégrée dans toutes les autres étapes du cycle, ce qui ouvrira des possibilités permanentes de reconsidération et de réévaluation des politiques (Höchtel et coll., 2016). Cette rétroaction continue permet de surveiller le progrès des programmes publics et de fournir, au fur et à mesure, des solutions pour améliorer les processus (Höchtel et coll., 2016; Pencheva et coll., 2020; Valle-Cruz et coll., 2020).

Grâce à l'analyse de données ainsi qu'aux capacités de simulation de l'IA, il devient possible de donner un retour d'information rapide et précis et de créer des scénarios pour chaque étape du cycle des politiques

publiques en temps quasi réel. Ces informations seraient particulièrement utiles pour les décideurs et elles contribueront à la redéfinition des problèmes sociétaux et à la formulation de nouvelles stratégies pour y répondre. Plus concrètement, la collecte et l'analyse de données provenant de sources variées, dont les réseaux sociaux, seraient nécessaires pour pouvoir fixer l'ordre du jour, puis formuler plusieurs options politiques et aider les acteurs à prendre des décisions plus éclairées rapidement. Après l'étape de la mise en œuvre, les résultats de l'évaluation permettront de déclencher une nouvelle itération de la politique publique. Ainsi, le cycle des politiques publiques, qui évoluait auparavant de manière linéaire, se produira de façon plus agile et incrémentale, grâce à l'utilisation des systèmes d'IA (Valle-Cruz et coll., 2020). Pour désigner ces transformations découlant de l'utilisation de l'IA, les études parlent alors du « cycle dynamique des politiques publiques » (*Dynamic Public Policy Cycle – DPPC*).

Les articles analysés montrent que la notion d'évaluation continue des politiques publiques basée sur l'IA offrirait un potentiel important aux organisations publiques. Cependant, celles-ci n'appliquent pas encore cette méthode, sauf dans le cas de quelques tentatives préliminaires. Par conséquent, on dispose de peu de données empiriques pour pouvoir juger de son efficacité et de l'étendue des défis qu'elle représente. La discussion autour de ses avantages potentiels et de ses obstacles demeure pour l'instant principalement théorique (Höchtel et coll., 2016).

Cette approche émergente permettrait de surveiller les politiques publiques tout au long du cycle et d'y apporter des modifications en temps réel en vue de les adapter rapidement aux changements et à l'évolution des besoins. Le recours à l'IA aurait aussi le potentiel de rendre le cycle des politiques publiques novateur, prédictif, précis, évolutif et ajustable (Valle-Cruz et coll., 2020).

L'évaluation continue à chaque étape du cycle des politiques réduirait les erreurs et les inefficacités dans l'élaboration des politiques en donnant la possibilité de les ajuster rapidement grâce à l'élaboration de scénarios prédictifs (Höchtel et coll., 2017; Valle-Cruz et coll., 2020). Elle ouvrirait également la porte à une plus large inclusion du public aux différentes étapes du cycle, ce qui est susceptible de favoriser non seulement la communication bidirectionnelle entre les citoyens et le gouvernement, mais aussi la relation de confiance du public (Höchtel et coll., 2016; Craglia et coll., 2020).

Selon les auteurs, outre les politiques plus souples et adaptatives, la mise en œuvre de l'IA dans les différentes étapes du cycle des politiques publiques pourrait conduire à une intervention politique plus ciblée, voire « personnalisée », en fonction des besoins et de la localisation des citoyens. Avec les données administratives de plus en plus fines et précises, il est possible d'identifier les caractéristiques clés, telles que l'âge, le sexe, le lieu de résidence et le profil socio-économique, et de commencer à mesurer l'impact des politiques publiques sur différents groupes, allant au-delà de la corrélation vers la causalité (Craglia et coll., 2020). Les analyses avancées offrent la possibilité d'observer simultanément des variables individuelles et agrégées (concernant par exemple le quartier, la ville, etc.) (Pencheva et coll., 2020). L'identification des points communs entre les groupes de citoyens grâce aux algorithmes aiderait à déterminer les « communautés » de besoins et d'intérêts partagés, pour lesquelles il serait possible d'adapter les politiques (Craglia et coll., 2020).

3. Les défis de l'IA en évaluation de programme

3.1. Les défis de l'évaluation continue des politiques publiques

L'utilisation de l'IA représente certes une occasion stratégique pour l'amélioration du processus d'élaboration des politiques publiques et leur efficacité. Cependant, modifier l'approche traditionnelle du cycle des politiques publiques et adopter les changements organisationnels et culturels nécessaires constitue tout un défi pour les organisations publiques (Valle-Cruz et coll., 2020). Pour y faire face, celles-ci doivent adopter une attitude d'ouverture et de collaboration pour permettre l'agrégation des données au-delà des frontières et des cloisons des départements administratifs. Elles sont aussi appelées à intégrer une certaine rapidité dans la prise de décision et l'exécution des tâches (Höchtel et coll., 2016). Le déploiement de l'IA doit tenir compte des capacités de l'organisation et de son ouverture au changement (Patel et coll., 2021).

Certains voient aussi dans l'usage de l'IA un risque de déshumanisation des politiques et une dépendance envers ces outils dans la prise de décision. Or, comme le mentionnent plusieurs auteurs, les données massives ne sont pas le remède qui permet de résoudre l'ensemble des problèmes relatifs à la formulation et à la mise en œuvre des politiques publiques. Se concentrer uniquement sur les processus d'évaluation conduit à résoudre des problèmes immédiats, mais présente le risque de passer à côté des causes sous-jacentes des problèmes sociaux que les programmes publics ont l'ambition de corriger. De plus, les données sur les décisions politiques ne sont souvent pas liées aux données sur la mise en œuvre des politiques, ce qui rend difficile d'effectuer une évaluation globale d'une politique publique (Pencheva et coll., 2020).

3.2. L'usage responsable et éthique des données

De manière générale, l'un des principaux risques associés à l'introduction de l'IA aux fins d'évaluation de programme concerne la gestion des données personnelles. En effet, les technologies de l'IA s'appuient sur des processus d'apprentissage qui doivent être nourris par de grands volumes de données. Les nouvelles possibilités pour l'IA dans le secteur public découlent justement de la capacité des institutions publiques à produire et collecter des données. En même temps, la gestion de ces données peut constituer une source de problèmes (Valle-Cruz et coll., 2020). Les organisations doivent d'abord veiller au respect des lois et règlements en vigueur concernant la confidentialité et la protection des données personnelles. Le nettoyage et l'anonymisation des données deviennent alors des étapes cruciales, même si cela peut être coûteux et chronophage (Pencheva et coll., 2020). L'équilibre entre les utilisations bénéfiques des données massives et les atteintes potentielles à la vie privée est difficile à établir (Höchtel et coll., 2016).

En outre, le recours aux systèmes d'IA pourrait soulever des préoccupations éthiques. L'IA pourrait intentionnellement ou accidentellement accroître les inégalités sociales et économiques, car une conception biaisée des algorithmes risque d'aboutir à une décision erronée. En règle générale, les modèles d'apprentissage automatique reflètent les données sur lesquelles ils ont été formés. S'il existe un biais dans les données de formation, le modèle reflétera celui-ci (Jasper et coll., 2019). La qualité des données et leur représentativité s'avèrent ainsi déterminantes pour éviter les effets discriminatoires des décisions prises par l'IA et garantir la qualité de l'évaluation.

La conception des systèmes d'IA doit être réalisée de manière transparente et responsable pour favoriser leur acceptation auprès des usagers et pour rassurer les sceptiques sur le fait que l'IA est sûre, responsable et efficace (Valle-Cruz et coll., 2020; Patel et coll., 2021).

3.3. La transformation du métier et des compétences

Comme pour plusieurs milieux professionnels, les technologies de l'IA sont susceptibles de transformer les métiers associés à l'évaluation de programme. L'IA automatise les tâches répétitives, fastidieuses et à faible valeur ajoutée qui étaient jusque-là effectuées par les humains. Alors que certains craignent le remplacement des fonctionnaires par des robots et des machines intelligentes, d'autres estiment que les changements dans les processus de travail induits par l'IA permettent aux professionnels de se concentrer davantage sur les tâches axées sur la créativité, l'innovation et la planification stratégique. Plus encore, avec la capacité d'analyser les données massives et de trouver des modèles et des informations précises, l'IA introduit dans le domaine de l'évaluation de nouvelles possibilités et des fonctionnalités que les humains ne pourront jamais réaliser (Valle-Cruz et coll., 2020).

Pour favoriser ces innovations dans le secteur public, les gouvernements doivent investir dans la formation des évaluateurs et des évaluatrices (Valle-Cruz et coll., 2020) qui ont désormais besoin de nouvelles compétences pour s'adapter à leurs nouvelles conditions de travail. Posséder des compétences en informatique et en analyse de données devient une exigence pour comprendre le fonctionnement des outils d'IA et accomplir efficacement leurs tâches (Beitel, 2005).

Conclusion

L'usage de l'IA aux fins d'évaluation de programme au sein des ministères et des organismes publics se limite encore à quelques expériences exploratoires. Cependant, la littérature indique que ces technologies sont susceptibles de transformer les pratiques d'évaluation de manière radicale. En effet, de nombreux chercheurs ont examiné les applications possibles de l'IA dans le domaine de l'évaluation, la faisabilité et l'efficacité de ces méthodes ainsi que les avantages et les défis qu'elles pourraient entraîner.

Ces études révèlent que les approches de NLP peuvent être appliquées dans différentes étapes du processus d'évaluation, à savoir la délimitation du périmètre de l'évaluation, la sélection des documents à inclure dans l'examen des projets identifiés et le codage et l'analyse des documents (Jasper et coll., 2019). De même, l'utilisation d'outils d'exploration de données d'IA permettrait d'approfondir le niveau d'analyse des données d'évaluation en répondant à des questions complexes concernant l'impact des programmes et leur efficacité (Beitel, 2005). Les études montrent aussi qu'il est possible de recourir aux techniques d'IA pour l'analyse avancée du contenu (Frazen et coll., 2022) et d'appliquer, plus spécifiquement, les méthodes d'apprentissage automatique pour l'étude des inférences causales en contexte d'évaluation de programme (Alcaud, 2021).

Les résultats préliminaires de ces travaux exploratoires sont prometteurs. Compte tenu de la vitesse avec laquelle ces technologies évoluent, les auteurs sont optimistes quant à l'avenir de l'utilisation de l'IA dans le secteur de l'évaluation. Ils s'accordent sur la nécessité de multiplier les expériences pour déterminer les avantages potentiels de l'IA sur les pratiques d'évaluation et en apprendre davantage sur les défis à surmonter (Frazen, 2022).

Grâce aux capacités de simulation de l'IA combinées aux capacités de traitement quasi instantané des vastes volumes de données, ces technologies favorisent l'évaluation continue dans toutes les phases du cycle des politiques publiques. La disponibilité de données rétroactives rapides et précises permettrait aux décideurs de suivre l'avancement des politiques et de les ajuster, au fur et à mesure, de manière à répondre aux besoins émergents, tout en se référant aux simulations fournies par les outils d'IA.

Parallèlement aux nombreux avantages que les organisations publiques pourraient tirer de l'introduction de l'évaluation basée sur l'IA, certains risques et défis sont à prévoir. L'adoption de l'évaluation continue des politiques publiques implique des changements organisationnels et culturels importants. De plus, les organisations doivent veiller au respect des lois et des règlements en vigueur concernant la confidentialité et la protection des données personnelles. Les algorithmes d'IA doivent être conçus de manière éthique, responsable et transparente pour éviter de perpétuer les inégalités sociales et économiques dans les décisions fournies par ces systèmes. L'autre défi qui s'impose aux organisations publiques consiste à préparer le personnel à l'usage de ces technologies pour une visée d'évaluation et à lui offrir les formations lui permettant d'acquérir les compétences nécessaires pour comprendre leur fonctionnement et les utiliser de manière optimale.

Bibliographie

- Alcaud, D. (2021). *L'inférence causale, enjeu de la gestion de la crise sanitaire*. <https://www.alliancy.fr/inference-causale-enjeu-gestion-crise-sanitaire>.
- Athey, S. et Imbens, G. W. (2017). The state of applied econometrics: causality and policy evaluation. *Journal of Economic Perspectives*, 31(2), 3-32.
- Beitel, S. E. (2005). *Applying artificial intelligence data mining tools to the challenges of program evaluation*. Thèse de doctorat, University of Connecticut.
- Craglia, M., Hradec, J. et Troussard, X. (2020). The big data and artificial intelligence: Opportunities and challenges to modernise the policy cycle. Dans V. Šucha et M. Sienkiewicz (dir.), *Science for Policy Handbook* (p. 97-104). Elsevier.
- Franzen, S., Quang, C., Schweizer, L., Budzier, A., Gold, J., Vellez, M., Ramirez, S. et Raimondo, E. (2022). *Advanced content analysis: Can artificial intelligence accelerate theory-driven complex program evaluation?* Independent Evaluation Group, World Bank Group.
- Höchtel, J., Parycek, P., Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, 26(1-2), 147-169. <http://dx.doi.org/10.1080/10919392.2015.1125187>.
- Jasper, P., Haldrup, S. V. et Mikhaylov, S. J. (2019). *Artificial intelligence and machine learning methods for programme evaluations in Global Affairs Canada*. Oxford Policy Management.
- Knaus, M. C. (2022). Double machine learning-based programme evaluation under unconfoundedness. *The Econometrics Journal*, 25(3), 602-627. <https://doi.org/10.1093/ectj/utac015>.
- Kreif, N. et DiazOrdaz, K. (2019). Machine learning in policy evaluation: New tools for causal inference. *Oxford Research Encyclopedias of Economics and Finance*. <https://doi.org/10.1093/acrefore/9780190625979.013.256>.
- Mohan, D., Harrison Bashingwa, J. J., Dane, P., Chamberlain, S., Tiffin, N. et Lefevre, A. (2019). Use of big data and machine learning methods in the monitoring and evaluation of digital health programs in India: An exploratory protocol. *JMIR Research Protocols*, 8(5). <https://doi.org/10.2196/11456>.
- Patel, J., Manetti, M., Mendelsohn, M., Mills, S., Felden, F., Littig, L. et Rocha, M. (2021). *AI brings science to the art of policymaking*. Boston Consulting Group. <https://www.bcg.com/fr-ca/publications/2021/how-artificial-intelligence-can-shape-policy-making>.

Pencheva, I., Esteve, M. et Mikhaylov, S. J. (2020). Big data and AI – A transformational shift for government: So, what next for research? *Public Policy and Administration*, 35(1), 24-44. <https://doi.org/10.1177/0952076718780537>.

Secrétariat du Conseil de trésor (SCT) (s.d.). *Évaluation de programme*. <https://www.tresor.gouv.qc.ca/performance-de-letat/cadre-de-gestion-de-ladministration-gouvernementale/evaluation-de-programme>.

Valle-Cruz, D., Criado, J. I., Sandoval-Almazán, R. et Ruvalcaba-Gomez, E. A. (2020). Assessing the public policy-cycle framework in the age of artificial intelligence: From agenda-setting to policy evaluation. *Government Information Quarterly*, 37(4). <https://doi.org/10.1016/j.giq.2020.101509>.