

INTELLIGENCE ARTIFICIELLE ET TRANSFORMATION DES MÉTIERS DE LA GESTION DOCUMENTAIRE

Steve Jacob

Seima Souissi

Charlie Martineau



UNIVERSITÉ
LAVAL

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

En partenariat avec :



Steve Jacob, Seima Souissi et Charlie Martineau

Intelligence artificielle et transformation des métiers de la gestion documentaire

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

Université Laval

Québec, 2022

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.

Charlie Martineau est étudiante à la maîtrise en études internationales à l'Université Laval. Elle est auxiliaire de recherche à la Chaire de recherche sur l'administration publique à l'ère numérique et membre étudiante du Centre d'analyse des politiques publiques (CAPP).

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 pose 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 de l'IA dans le domaine de la gestion documentaire.....	3
1.1. Extraire des données et indexer des documents	4
1.2. Évaluer et classifier des documents	5
1.3. Examiner la sensibilité des données	6
1.4. Gérer le courrier électronique	7
2. Transformations du rôle et des compétences des professionnels de la documentation.....	8
2.1. Un rôle à redéfinir au sein d'une gouvernance intégrée	8
2.2. Transformation des compétences	9
3. Défis et obstacles à l'utilisation de l'IA dans le domaine de la gestion documentaire.....	10
Conclusion	12
Bibliographie	13

Introduction

À travers l'organisation de l'information, la gestion documentaire vise à faciliter l'accès aux documents, leur consultation, leur diffusion et l'archivage de ceux-ci et des informations qu'ils contiennent. Elle regroupe l'ensemble des procédures et des moyens permettant d'organiser l'information documentaire tout au long du cycle de vie des documents (SCT, 2004). Il s'agit des procédures prescrites ou recommandées pour régir l'utilisation et la conservation des documents papier et électroniques de natures et de sources multiples.

Depuis plusieurs années, l'environnement de la gestion documentaire connaît des transformations importantes. Premièrement, l'arrivée du numérique a profondément transformé la gestion documentaire et a entraîné un changement de paradigme. La gestion, qui était centrée sur le cycle de vie documentaire, évolue maintenant vers une gestion des flux d'information (Caron, 2021). Deuxièmement, l'accès à l'information n'est plus nécessairement lié à l'accès aux documents (Lamouroux et Ferchaud, 2006). Troisièmement, à partir des années 1990, la numérisation à grande échelle dans les administrations publiques et privées a transformé de nombreuses collections en données numériques. Quatrièmement, avec l'explosion des documents numérisés et numériques à travers l'Internet des objets et les médias sociaux, le processus de gestion documentaire, basé sur le cycle de vie, à savoir l'évaluation, la sélection et l'examen des documents par les professionnels humains, est de moins en moins possible. Des machines ont été introduites progressivement pour prêter main-forte aux employés dans la gestion des grands volumes de données disponibles. Les archives deviennent peu à peu une organisation de mégadonnées qui s'appuie largement sur les capacités de technologies dotées d'une intelligence artificielle (IA) pour gérer les flux d'information. L'IA pourrait être utilisée tout au long des processus de gestion documentaire et changer radicalement la manière dont les documents sont créés, traités, stockés et partagés. Ces technologies seraient appliquées à la fois pour faire évoluer les activités traditionnelles relatives à la gestion de documents et pour expérimenter de nouvelles façons de répertorier et d'organiser les informations et d'y accéder (Colavizza et coll., 2022).

Le recours à l'IA dans le domaine de l'information et documentation va modifier les conditions d'exercice de plusieurs métiers, comme ceux de bibliothécaire, d'archiviste, de documentaliste et de catalogueur. Qu'ils travaillent dans le secteur privé ou public, dans les bibliothèques ou autres centres de documentation, ces professionnels de la gestion documentaire sont appelés à maîtriser de nouveaux outils et à développer de nouvelles compétences pour répondre aux nouvelles exigences de leur métier et remplacer les procédures et normes professionnelles qui deviennent obsolètes (Utard, 2017).

Dans le présent rapport, nous explorons d'abord certaines applications de l'IA dans le secteur de la gestion documentaire et leurs répercussions sur le fonctionnement des archives et sur la gestion des données au sein des organisations ainsi que sur le rôle et les compétences des professionnels de la documentation. Nous nous attardons ensuite sur les principaux avantages et défis associés à l'implantation de ces technologies dans le domaine.

1. Usages de l'IA dans le domaine de la gestion documentaire

Le recours aux technologies de l'IA a le potentiel d'améliorer l'efficacité des processus de gestion documentaire. Les opérations pourraient être réalisées plus rapidement, sur une volumétrie plus grande et avec un risque d'erreur amoindri. Pour les organisations qui recourent à ces technologies, les factures, les documents liés aux ressources humaines (RH) et les contrats figurent en tête des éléments ciblés par la gestion automatisée, en raison de leur volume important et de la tâche chronophage que représentent les étapes de traitement relatives à la gestion de ces documents. L'IA permet à l'organisation de réaliser des gains significatifs en matière de productivité et, par conséquent, de rentabiliser rapidement les sommes investies dans l'implantation de ces technologies (Bergonzoli, 2018). De plus, au fil des années, les coûts relatifs à cet investissement diminuent progressivement à mesure qu'augmente la concurrence entre les entreprises émergentes (*start-up*) et les fournisseurs de services technologiques offrant des prix compétitifs et adaptés au marché (Rolan et coll., 2019).

Le principal avantage associé à l'intégration de l'IA dans les plateformes de gestion documentaire consiste à valoriser les informations « dormantes » ou abandonnées. Généralement, les organisations possèdent une quantité importante de documents en différents formats et souvent dispersés sur différents supports (messagerie, outils de stockage et de synchronisation dans l'informatique [*cloud*], etc.). Avec l'IA, les organisations peuvent se réappropriier ce corpus et passer d'un traitement documentaire cloisonné à une gouvernance d'un capital informationnel transversal (Bergonzoli, 2018). L'indexation et la classification des documents à l'aide des technologies de l'IA fourniraient des références qu'il est plus facile d'exploiter grâce aux métadonnées, c'est-à-dire des informations descriptives, au sein de fragments de textes porteurs d'une information précise et recherchée. L'analyse sémantique des contenus faciliterait le repérage et l'intégration des données et ferait gagner du temps aux utilisateurs (Grandmontagne, 2018).

En effet, le recours à l'IA pour la gestion documentaire permettrait de répondre adéquatement aux besoins d'un plus grand nombre d'utilisateurs qui ont des attentes variées. Par exemple, pour les chercheurs et les universitaires, ces technologies pourraient améliorer l'accès aux données et enrichir la façon dont ils peuvent les utiliser. Dans le contexte organisationnel, la littérature rapporte l'exemple d'une entreprise employant plus de 100 000 salariés. Celle-ci utilise un outil de gestion documentaire pour que les ingénieurs nouvellement recrutés puissent accéder aux informations sur le fonctionnement des vieux équipements toujours en utilisation, mais pour lesquels l'entreprise n'investit plus dans la formation. L'outil doté d'une IA fait le tri parmi les informations contenues dans les milliers de documents rédigés par les anciens employés, puis il repère uniquement celles dont les ingénieurs actuels ont besoin (Ferguson, 2019).

Par ailleurs, des experts estiment que l'IA a le potentiel de favoriser une forme plus inclusive et plus équitable de l'archivage. Les normes de sélection et de description automatisées des archives sont susceptibles de faire apparaître, de préserver et de mettre à la disposition des utilisateurs des références appartenant à des communautés sous-représentées et des groupes minoritaires. L'IA pourrait détecter des sources documentaires sur Internet et sélectionner le contenu pertinent provenant de sites développés par des groupes ou organisations minoritaires. Cette application de l'IA comme outil d'archivage et de gestion documentaire pourrait favoriser la prise en compte d'une diversité de sources et contribuer à une pratique sociotechnique éthique et inclusive (Colavizza et coll., 2022).

Les applications et usages de l'IA dans le champ de la gestion documentaire se multiplient. Dans les paragraphes suivants, nous en illustrons quelques-uns en nous attachant sur les usages de l'IA pour des fonctions spécifiques du processus de gestion documentaire. Ces fonctions sont : 1) l'extraction des données et l'indexation des documents, 2) l'évaluation et la classification des documents, 3) l'examen de la sensibilité des données ainsi que 4) la gestion du courrier électronique.

1.1. Extraire des données et indexer des documents

L'utilisation de l'IA dans l'organisation des archives commence par **l'extraction automatique des contenus et leur indexation**, c'est-à-dire la représentation du contenu d'un document par des expressions linguistiques. En effet, l'IA permet d'aller au-delà des principes traditionnels d'organisation des archives et de leur accès (provenance, nom de fichier, date de création, type de contenu et auteur). Cette approche peut être complétée par une indexation axée sur le contenu des documents (Colavizza et coll., 2022). Cette indexation consiste en l'ajout d'annotations ou de métadonnées qui leur sont associées. **L'objectif est de dégager les éléments qui sont évoqués dans les textes rédigés en langue naturelle**, de manière à donner un sens à des documents d'archives de plus en plus volumineux (Hubain et coll., 2016). Cette annotation peut être une information (reconnaissance d'entités nommées) ou bien un lien vers une source d'information (désambiguïsation d'entités nommées).

L'extraction automatisée d'information à partir de ressources textuelles est appelée « **forage de textes** » (*text mining*). Cette pratique facilite l'indexation des documents et la conception de taxonomie. Elle s'appuie sur des méthodes et des outils relevant du domaine du traitement automatique du langage naturel (*natural language processing – NLP*) et qui permettent de faire correspondre les chaînes de caractères des documents analysés avec les termes et les concepts définis dans la taxonomie. Deux méthodes d'extraction d'information sont possibles : l'extraction de termes et l'extraction de relations sémantiques entre ces termes. L'extraction de relations sémantiques a pour objectif de structurer la taxonomie et consiste – dans le cas d'un thésaurus – dans la détection des relations sémantiques et de contiguïté entre les termes (Hubain et coll., 2016). Des logiciels libres (*open source*) de fouille de textes sont disponibles et sont accessibles à une communauté importante d'utilisateurs et de développeurs (Hubain et coll., 2016).

De plus, les solutions de gestion de contenu de nouvelle génération, appelées « plateformes de services de contenu », comme Nuxeo Insight, proposent des modèles de métadonnées flexibles et extensibles qui permettent de peaufiner l'indexation et d'ajouter de nouveaux champs au besoin. Cette fonction est considérée comme une avancée par rapport aux systèmes précédents dans lesquels, une fois les métadonnées définies et stockées, il était difficile d'introduire des changements sans effectuer une mise à jour de l'ensemble du contenu associé à ces métadonnées. Les documents d'archives peuvent désormais être reconnus par un plus grand nombre de métadonnées de manière à apporter de la valeur, en tenant compte du contexte et de la connaissance de l'écosystème de gestion de l'information.

Les plateformes de services de contenu dotées d'une IA offrent aussi l'avantage de connecter les contenus qui se trouvent au sein des différents systèmes de gestion de l'information. Elles permettent ainsi de récupérer les références stockées dans les solutions d'ancienne génération et d'offrir une vision centralisée de l'information via une interface unique (Jones, 2018).

1.2. Évaluer et classifier des documents

Le rôle de l'IA dans l'évaluation documentaire est décrit dans la littérature portant sur l'automatisation des processus d'archivage. Les professionnels de la gestion documentaire sont aux prises avec des volumes sans cesse croissants de documents non structurés et non catégorisés. En combinant les techniques de traitement du langage naturel et l'apprentissage automatique¹, les solutions d'IA ont le potentiel de faciliter l'évaluation de grandes quantités de documents.

Le recensement et l'extraction de données pertinentes permettent de reconnaître les documents qui ont une valeur archivistique pour l'organisation et ceux qui peuvent être éliminés (Jones, 2018). Souvent, les organisations disposent de réglementations et de politiques de conservation qui leur sont propres et qui les obligent à conserver certains types de documents pour une durée bien précise. Cependant, sans une connaissance fine du contenu disponible, il est impossible de savoir quand un document doit être détruit. Le tri manuel est non seulement fastidieux et chronophage, mais aussi sujet aux erreurs. Les outils basés sur l'IA sont susceptibles de rendre le classement des documents plus facile et plus rapide (Jones, 2018).

Les études montrent que le contenu disponible au sein d'une organisation peut être redondant, obsolète et trivial. Éliminer ce contenu à l'aide de l'IA facilite, par la suite, la détermination des ressources pertinentes auxquelles seront appliquées les politiques de conservation. Ces outils permettraient plus précisément d'analyser le contenu restant de façon détaillée, de lui associer des règles de conservation et de formuler les recommandations appropriées quant à son élimination ou à son archivage. Cette **suppression du contenu superflu** rendrait les processus de recensement, de déclaration et de gestion des archives plus pertinents et moins coûteux (Jones, 2018).

Cette fonction est utilisée par exemple dans le secteur bancaire. Désormais, il est possible de sélectionner automatiquement les documents des personnes qui ne sont plus clientes depuis un certain temps, de purger le fonds documentaire (format papier ou électronique), d'effectuer une désambiguïsation des contenus², puis de classer les documents et d'en ressortir les métadonnées essentielles (Archimag.com, 2021).

Les outils d'IA qui effectuent la classification des documents et qui fonctionnent à l'aide d'**algorithmes d'apprentissage automatique** sont entraînés, au préalable, sur un corpus de documents ayant été étiquetés manuellement avec leur classification correcte. Ces algorithmes peuvent apprendre à reconnaître une combinaison de caractéristiques représentant « l'empreinte digitale » d'une catégorie donnée et à faire des prédictions sur la classification des nouveaux documents (Hutchinson, 2020).

¹ Le traitement du langage naturel est la capacité pour un programme informatique de comprendre le langage humain tel qu'il est parlé (TechTarget, 2018). L'apprentissage automatique (*machine learning*) fait référence au processus par lequel les ordinateurs développent la capacité d'apprendre et de faire des prévisions à partir de données, puis d'effectuer des ajustements, plutôt que de suivre des instructions explicitement programmées pour le faire (Data Franca, 2022).

² La désambiguïsation lexicale permet de déterminer le sens d'un mot dans une phrase lorsque celui-ci peut avoir plusieurs sens. Par exemple, le mot orange peut renvoyer à une couleur, à un fruit, à un opérateur de téléphonie ou encore à une ville française.

Les outils d'IA atteignent déjà un bon niveau de précision pour faire correspondre automatiquement les nouveaux corpus aux classes appropriées (Rolan et coll., 2019; Colavizza et coll., 2022).

La classification des documents à l'aide des outils d'IA va au-delà de l'évaluation visant à conserver ou à détruire un enregistrement. Elle peut inclure d'autres classifications binaires permettant notamment de déterminer si un courriel est ou n'est pas un pourriel (*spam*). **La classification à l'aide de l'IA peut impliquer de nombreuses classes** (1 sur N) ou des classes à valeurs multiples (par exemple, le document porte sur le sujet X, l'image contient les éléments A, B et C). L'algorithme procède par le repérage, dans les documents, d'éléments d'information appelés « caractéristiques ». Dans l'analyse de documents textuels, les caractéristiques renvoient à l'existence et aux emplacements de mots, de phrases ou d'autres éléments grammaticaux particuliers ainsi qu'à des indices de métadonnées. Dans le traitement d'images, les caractéristiques concernent les propriétés des pixels (couleur, luminosité, niveau de granularité, contours, formes, etc.) (Rolan et coll., 2019).

Cette capacité de l'IA à faire ressortir l'information pertinente et à la départager peut être utilisée par les organisations pour déterminer la nature des documents qu'elles possèdent et pour les classer en séparant, par exemple, les contrats, les factures, les brochures. Il serait aussi possible d'alimenter les activités du département du service à la clientèle en rassemblant tous les contrats concernant un client particulier. Ces requêtes ne pourraient être exécutées si le contenu archivé n'était pas doté des métadonnées nécessaires aux recherches (Jones, 2018).

Grâce à ces outils d'IA, les organisations peuvent classer les documents avec précision, masquer automatiquement les données à caractère sensible et créer une base de connaissances. De plus, l'IA permet aussi de détecter certaines tendances dans les informations disponibles. Par exemple, le traitement automatique du courriel d'une organisation pourrait l'aider à savoir, en amont, si un client est insatisfait des services ou est en passe de résilier son contrat (Archimag.com, 2021).

1.3. Examiner la sensibilité des données

Plusieurs auteurs se sont penchés sur l'application du traitement du langage naturel et de l'apprentissage automatique à l'examen de sensibilité, c'est-à-dire pour la détection des informations personnelles dans les documents. Étant soumises à une législation stricte sur la protection des renseignements personnels, les organisations gouvernementales sont particulièrement concernées par ce défi (Hutchinson, 2020). Elles doivent empêcher la divulgation d'informations sensibles ou de renseignements personnels, tout en assurant l'accessibilité des données aux demandeurs. Les technologies de l'IA offrent des solutions diverses pour filtrer les données et détecter le contenu sensible. Une approche de classement automatique permettra ensuite de séparer les contenus qui peuvent être divulgués par l'organisation de ceux qui doivent demeurer confidentiels (Colavizza et coll., 2022).

Des experts estiment toutefois que la notion de sensibilité est difficile à cerner précisément, car le sujet d'un document n'est pas toujours un élément suffisant pour déterminer que les données qu'il contient sont sensibles. En effet, la sensibilité dépend fortement des contextes dans lesquels les documents sont produits et utilisés. Par conséquent, il est important de recourir à des technologies avancées permettant d'intégrer l'analyse contextuelle à cette opération (Colavizza et coll., 2022).

1.4. Gérer le courrier électronique

Les capacités de l'IA dans la gestion des données structurées et non structurées ainsi que dans l'indexation, la classification et l'examen de sensibilité des documents pourraient être exploitées pour gérer le courrier électronique.

Pour donner un exemple concret des outils disponibles, mentionnons ePADD, un logiciel développé par les bibliothèques de l'Université de Stanford pour la gestion des archives de courrier électronique. Il combine les techniques de traitement du langage naturel, la linguistique computationnelle et l'apprentissage automatique. Initialement destiné aux organisations ayant une fonction mémorielle comme les bibliothèques, les archives ou les musées pour gérer le courrier électronique à valeur historique et culturelle, ePADD est aujourd'hui utilisé par différentes organisations disposant d'un grand volume de courrier électronique. En effet, le courrier électronique joue un rôle crucial dans le fonctionnement des organisations modernes, mais l'important volume et la nature non structurée des courriels rendent difficiles leur gestion, leur élimination et leur examen de sensibilité (Rolan et coll., 2019). Le logiciel ePADD peut réaliser de nombreuses tâches comme l'extraction d'entités nommées et la désambiguïsation du contenu en vue de déterminer la nature des documents contenus dans les courriels (contrat, correspondance, facture, etc.) (Jones, 2018). Doté d'une fonction d'analyse lexicale qui permet le repérage des termes associés à des informations personnelles, l'outil peut également être utilisé pour filtrer les courriels contenant des informations sensibles, confidentielles ou légalement restreintes et de les classer conformément aux politiques et aux réglementations en vigueur (Hutchinson, 2020). L'ajout de mots-clés ou d'étiquettes (*tag*) assure aussi de mieux comprendre les contenus de courriels et d'en faciliter la recherche et l'accès.

2. Transformations du rôle et des compétences des professionnels de la documentation

2.1. Un rôle à redéfinir au sein d'une gouvernance intégrée

La gestion des ressources documentaires pourrait être bonifiée par des outils d'IA permettant de traiter un volume de plus en plus important de documents et de données. Selon les auteurs, ces technologies pourraient jouer un **rôle libérateur pour les professionnels de la documentation**, car les tâches répétitives et chronophages seraient prises en charge par les machines. Ces changements ne signifient pas que le rôle de ces professionnels est dépassé. Au contraire, leur intervention dans le processus de gestion documentaire automatisée est toujours requise et leur travail est en train d'évoluer vers de nouvelles responsabilités susceptibles d'apporter une plus grande valeur ajoutée à leur métier (Theimer, 2018).

L'IA contribuerait à favoriser une reconceptualisation des archives et de la documentation en gestion de données « qui ont du sens » et, par conséquent, un **repositionnement du rôle des professionnels du domaine de la gestion de supports à celui de la gestion de métadonnées**. Dans leurs activités, les professionnels de la gestion documentaire auront la possibilité de se concentrer davantage sur l'analyse du contenu des données et leur structuration, et non plus uniquement sur leur gestion et leur conservation. Avec la massification des données, ces tâches seront réalisées en collaboration avec des informaticiens et des gestionnaires de centres de données (Servais et Mirguet, 2015). Les professionnels de la gestion documentaire seraient ainsi appelés à orienter leurs activités vers la **création de sens, la narration et la fourniture de contexte autour des données** pour favoriser leur exploitation et leur cheminement au sein de l'organisation et fournir une valeur ajoutée aux utilisateurs (Bunn, 2018; Colavizza et coll., 2022).

Malgré l'automatisation de très nombreuses tâches, l'intervention humaine dans le processus de la gestion documentaire demeure indispensable pour : alimenter et vérifier la qualité des algorithmes, valider une décision prise par un algorithme, résoudre des problèmes ou encore traiter les cas difficiles. **L'automatisation du processus de la gestion documentaire exige toujours un accompagnement et un suivi par les professionnels humains**, de manière à garantir la qualité du travail et la confiance des utilisateurs (Colavizza et coll., 2022).

Dans un contexte où l'abondance d'information coexiste avec l'absence d'information de qualité, la gestion de l'information est devenue un enjeu de société majeur. Les spécialistes de la gestion documentaire devraient ainsi veiller à garantir l'intégrité, la fiabilité et l'authenticité des informations, malgré un environnement particulièrement propice à la désinformation et à la manipulation de l'information (Brown, 2018). Comme pour d'autres activités professionnelles, les métiers de la gestion documentaire reposent sur des valeurs et des principes éthiques qu'il convient de préserver, c'est pourquoi la touche humaine demeure nécessaire (Theimer, 2018; Bergonzoli, 2018). Par conséquent, les professionnels du domaine de la gestion documentaire devront adapter leurs pratiques au nouvel environnement, tout en maintenant un regard critique sur les répercussions des changements technologiques, notamment en matière de transparence, d'égalité et de respect des renseignements personnels (Schallier, 2019).

Avec la multiplication des unités qui se partagent la gestion de l'information au sein des organisations modernes, les professionnels de l'information devraient s'affirmer comme des acteurs essentiels dans la gouvernance de l'information de leur organisation et s'impliquer dans les projets et les comités de haut niveau consacrés à ce sujet. Pour ce faire, ils peuvent emprunter deux chemins. D'une part, les professionnels devraient acquérir et mobiliser des compétences en lien avec le pouvoir organisationnel, c'est-à-dire une certaine aptitude à influencer et à convaincre les autres (Zwarich et Maurel, 2020). D'autre part, ils devraient ajuster leurs pratiques afin de les accorder au nouvel environnement dans lequel ils évoluent, en raison du changement de paradigme centré sur les flux d'information.

2.2. Transformation des compétences

Avec l'introduction de l'IA dans le domaine de la gestion documentaire, les professionnels de ce secteur sont amenés à mettre à jour leurs compétences pour s'adapter aux nouvelles transformations. Dans les années à venir, les experts prévoient l'émergence de nouveaux métiers centrés sur l'exploitation des sciences des données et des technologies de l'IA (Lamouroux et Ferchaud, 2006).

Les professionnels de la gestion documentaire devraient suivre des **formations et, de préférence, détenir de l'expérience dans les techniques d'IA** afin de comprendre comment déployer cette technologie et en tirer profit dans les processus de documentation et d'archivage. Ces connaissances sont nécessaires pour démystifier la logique de fonctionnement de ces outils et évaluer leurs résultats, mais aussi pour être en mesure de collaborer au sein d'équipes multidisciplinaires. Comme lors des vagues technologiques antérieures qui ont eu un impact sur les métiers de la documentation, une telle reconversion serait nécessaire pour compenser l'écart de connaissances et de compétences qui se creuse entre les informaticiens et les experts en science des données impliqués dans les processus de la gestion documentaire et les professionnels de la gestion documentaire (Rolan et coll., 2019; Colavizza et coll., 2022).

Même si une bonne partie des opérations pourrait être confiée aux algorithmes et aux informaticiens qui les conçoivent, les professionnels de la gestion documentaire demeureront les **cogarants de la qualité des données et d'une utilisation appropriée de l'IA**, dans le respect des lois et des principes éthiques (Colavizza et coll., 2022). Une double compétence en analyse documentaire et en informatique serait alors indispensable pour leur permettre d'accomplir leur mission et d'accompagner le travail des informaticiens.

Ainsi, en plus de l'apprentissage des méthodes et des principes de la gestion documentaire traditionnelle, les futurs professionnels du secteur devraient suivre de la formation dans d'autres disciplines comme l'analyse informatique, la science des données, l'exploitation du Web sémantique et la maîtrise des métadonnées (Theimer, 2018; Hutchinson, 2020; Lavigueur, 2020).

Par exemple, considérant la transformation de la pratique archivistique par les techniques d'IA avancées telles que l'apprentissage automatique, des auteurs soutiennent la création d'une nouvelle discipline à l'intersection de l'archivistique et de l'informatique et suggèrent des changements profonds dans l'enseignement de l'archivistique. Cette discipline émergente appelée « archivistique computationnelle » vise à attirer les archivistes vers le raisonnement logique des ordinateurs, et les informaticiens et ingénieurs en logiciel vers les méthodes et les principes de l'archivistique (Colavizza et coll., 2022).

Par ailleurs, l'implantation des nouvelles technologies dans le domaine de la gestion documentaire permet aux professionnels de gagner du temps qu'ils peuvent consacrer à des tâches relevant davantage de la gestion de projets et de la communication. Pour y parvenir, les professionnels de la gestion documentaire devraient disposer des compétences managériales et communicationnelles nécessaires pour interagir avec une variété d'interlocuteurs travaillant dans les domaines de l'informatique, de l'animation culturelle et scientifique, de la communication et du droit, comme les ingénieurs et responsables de centres de données, les enseignants et chercheurs et les responsables de la protection des données personnelles (Servais et Mirguet, 2015). Cette transformation requiert également une adaptation de la gouvernance organisationnelle pour qu'elle favorise davantage la collaboration et le décloisonnement des unités. Des efforts en vue d'ériger cette gouvernance intégrée doivent encore être entrepris puisque « l'existence d'une gouvernance opérante à l'intérieur des institutions publiques qui permettrait de mieux comprendre et canaliser la production documentaire reste encore à inventer et à instaurer. Les regards portés sur les problématiques de la gestion de la production documentaire sont [encore] partiels et disciplinaires : accès, protection, technologies, archivistique, bibliothéconomie, etc. » (Caron, 2017, p. 860).

3. Défis et obstacles à l'utilisation de l'IA dans le domaine de la gestion documentaire

L'implantation de l'IA dans la gestion documentaire n'est pas sans soulever des défis pour les organisations. Ces défis ont trait principalement à la disponibilité des données pour la conception des outils, à la capacité de stockage et à la sécurité des renseignements personnels.

L'hétérogénéité des formats, des supports et des sources de données d'archivage entre les fichiers, les séries, les collections et les institutions constitue le premier obstacle à la généralisation des initiatives d'automatisation (Rolan et coll., 2019).

De plus, le développement de modèles d'apprentissage en profondeur pour l'évaluation et le classement des références **exige une grande quantité de données, dans un format adapté à ces technologies** dans le but d'atteindre un niveau satisfaisant de performance. Les experts précisent qu'environ 5 000 exemples étiquetés par catégorie seraient nécessaires pour entraîner un algorithme d'apprentissage profond. La précision du modèle augmente avec la multiplication du volume de données d'apprentissage. Elle pourrait égaler ou dépasser les performances humaines si l'algorithme est formé avec un ensemble de données contenant au moins 10 millions d'exemples étiquetés (Rolan et coll., 2019).

Ce processus est d'autant plus long et exigeant qu'une fois formés, les modèles d'apprentissage profond doivent être testés sur différentes données avant d'être appliqués en production.

Au besoin de grands volumes de données s'ajoute le besoin de ressources informatiques suffisantes, car la taille des flux de données et leur traitement via les technologies d'IA exigent une bande passante à capacité élevée et de grands espaces de stockage en infonuagique (*cloud*) (Grandmontagne, 2018).

À cet égard, mentionnons que les capacités de stockage numérique se sont développées de manière exponentielle et que leurs coûts ont nettement diminué au cours des dernières années. Cependant, la capacité à créer de nouvelles données dépasse rapidement les capacités de stockage disponibles. Malgré les progrès, il est difficile d'envisager que tout puisse être préservé. La fonction d'évaluation démontre alors toute sa pertinence. Il est essentiel que l'organisation soit en mesure de déterminer ses besoins documentaires, pour concevoir ensuite les systèmes d'information et d'archivage intelligents qui créent, capturent, décrivent, gèrent et préservent les documents (Brown, 2018).

L'autre défi associé à l'usage de l'IA, et plus particulièrement aux modèles d'apprentissage profond, réside dans l'opacité des algorithmes. Le nombre croissant de couches et de nœuds des systèmes d'apprentissage profond ainsi que leur interdépendance et leurs paramètres complexes aboutissent à des systèmes de « boîte noire » dont le fonctionnement est difficile à interpréter. Ni l'utilisateur ni le programmeur ne sont en mesure d'accéder au raisonnement conduisant à une décision ou à une recommandation produite par ces algorithmes (Rolan et coll., 2019; Ferguson, 2019).

Par conséquent, certains auteurs soulignent la difficulté d'appliquer cette technologie dans la prise de décisions automatisées touchant les dossiers complexes des individus. Dans certains contextes, la décision de conserver ou de détruire un document peut entraîner des conséquences importantes sur les personnes. L'implication de l'humain demeure ainsi nécessaire, car la familiarité avec le contenu des dossiers et la connaissance des contextes s'avèrent indispensables. De plus, les organisations ne doivent pas perdre de vue les risques de biais associés à ces technologies et les conséquences éthiques qu'ils peuvent entraîner (Rolan et coll., 2019).

Les technologies d'IA faciliteraient le déploiement de nouvelles méthodes de collecte de données et permettraient de collecter et de créer de nouvelles ressources nées numériques, à travers notamment les systèmes liés à l'Internet des objets (par exemple : données de capteurs de qualité de l'eau et reconnaissance automatique des plaques d'immatriculation). Les organisations utilisant de tels systèmes devraient s'assurer que la manière dont les informations sont collectées et organisées est conforme aux lois, réglementations et attentes concernant la protection des renseignements personnels. Les organisations gouvernementales devraient veiller à ce que les nouvelles formes d'archivage numérique respectent également les valeurs publiques comme la responsabilité, la transparence et l'intégrité (Colavizza et coll., 2022).

La littérature met aussi en évidence les risques liés à l'utilisation des services d'infonuagique (*cloud*) pour le stockage des documents, et particulièrement la confidentialité des renseignements personnels ainsi que la propriété légale des données traitées (Rolan et coll., 2019). Le recours à ces technologies doit s'accompagner d'une gestion rigoureuse des accès et trouver un équilibre entre une facilité d'accès qui risque de provoquer des problèmes de sécurité (Bergonzoli, 2018) et un réflexe de surprotection qui emprisonne les données et freine leur diffusion et leur utilisation (Grandmontagne, 2018).

Conclusion

Les technologies d'IA transforment les pratiques de conservation des documents physiques et électroniques. Elles ont le potentiel de faciliter l'accès et la diffusion des informations qu'ils contiennent. Ces technologies peuvent être mobilisées à toutes les étapes du processus de la gestion documentaire, de la création jusqu'à la consultation des données. Les fonctionnalités avancées offertes par l'IA en matière de gestion documentaire permettraient aux organisations de reprendre le contrôle de leurs données et de mieux les exploiter. Elles sont aussi susceptibles d'en améliorer l'accès et de répondre plus adéquatement aux besoins des utilisateurs.

Les technologies de traitement du langage naturel rendent possible une indexation des documents basée sur leur contenu. Elles permettent la détection d'éléments linguistiques dans les textes et, par conséquent, l'ajout de métadonnées susceptibles d'enrichir la sélection des documents. Cette extraction automatisée d'information faciliterait aussi la taxonomie des références, c'est-à-dire leur classement à l'aide des métadonnées répertoriées et leurs relations sémantiques. Grâce à cette technologie, les documents peuvent désormais être annotés avec un nombre plus important de métadonnées plus précises, ce qui contribue à contextualiser davantage les documents et à en améliorer leur traçabilité et leur repérage.

Conjuguées à la technologie de l'apprentissage automatique, les techniques de traitement du langage naturel faciliteraient l'évaluation de grandes quantités de documents et permettraient de distinguer ceux qui doivent être conservés de ceux qui peuvent être détruits. Elles peuvent servir aussi à déterminer la nature des documents et à les classer en conséquence. Cette capacité à extraire les informations dans les documents serait particulièrement utile pour l'examen de la sensibilité des documents, c'est-à-dire l'identification des informations personnelles en vue de départager les contenus qui peuvent être divulgués par l'organisation et ceux qui doivent rester confidentiels ou encore obéir à une réglementation particulière.

Dans ce nouveau contexte, le rôle des professionnels de l'information et de la documentation serait orienté vers la gestion des métadonnées, la création de sens et la contextualisation des données. Ces professionnels sont appelés à travailler en complémentarité avec les machines et à apprendre à tirer le meilleur profit de ces technologies. L'acquisition de nouvelles connaissances dans les domaines de l'informatique et de la science des données serait nécessaire pour comprendre de fonctionnement de ces outils et évaluer leurs résultats. Les experts estiment que le domaine de la gestion documentaire évolue vers une nouvelle discipline fusionnant les pratiques de l'archivistique et de la documentation avec les connaissances des sciences de l'informatique. De nouveaux métiers sont amenés à apparaître dans l'avenir.

Cependant, malgré le développement de leurs capacités analytiques et décisionnelles, les outils d'IA nécessitent encore un accompagnement de la part des spécialistes de la gestion documentaire. En effet, l'opacité des algorithmes notamment rend difficile l'interprétation des résultats et des conséquences que ces résultats pourraient entraîner sur les individus. Ces professionnels demeurent les cogarants des principes de fiabilité, de responsabilité et de transparence des données, mais aussi d'une utilisation éthique de l'IA dans le domaine de la gestion documentaire.

Bibliographie

- Archimag.com. (2021). *Xelians s'appuie sur l'IA pour mettre fin au vrac numérique*. <https://www.archimag.com/demat-cloud/2021/04/30/gestion-documentaire-intelligence-artificielle>
- Bergonzoli, F. (2018). L'automatisation au cœur de la gestion documentaire. *Solutions numériques*.
- Brown, C. (dir.). (2018). *Archival futures*. Londres, Facet Publishing.
- Bunn, J. (2018). Frames and the future of archival processing. Dans C. Brown (dir.), *Archival futures* (p. 65-78). Londres, Facet Publishing.
- Caron, D. J. (2017). La production documentaire dans les administrations publiques : enjeux et pistes de solution. Dans N. Michaud (éd.), *Secrets d'États? : les principes qui guident l'administration publique et ses enjeux contemporains* (p. 807-839). Québec, Presses de l'Université du Québec.
- Caron, D. J. (2021). *Écosystème de la transformation de l'administration publique vers le numérique*. Québec, Presses de l'Université du Québec.
- Colavizza, G., Blanke, T., Jeurgens, C. et Noordegraaf, J. (2022). Archives and AI: An overview of current debates and future perspectives, *Journal on Computing and Cultural Heritage*, 15(1), 1-15.
- Data Franca. (2022). *Apprentissage automatique*. https://datafranca.org/wiki/Apprentissage_automatique
- Ferguson, Y. (2019). Ce que l'intelligence artificielle fait de l'homme au travail. Visite sociologique d'une entreprise. Dans F. Dubet, *Les mutations du travail* (p. 23-42). Paris, La Découverte.
- Grandmontagne, Y. (2018). *L'IA dans la gestion documentaire*. IT Social. <https://itsocial.fr/enjeux-it/enjeux-innovation/intelligence-artificielle/lia-gestion-documentaire>
- Hubain, R., Hooland, S. et Verborgh, R. (2016). Classification automatisée : Rêve ou réalité? Analyse critique de l'usage du *text mining* pour la conception de vocabulaires contrôlés. *I2D - Information, données et documents*, 2(53), 70-79.
- Hutchinson, T. (2020). Natural language processing and machine learning as practical toolsets for archival processing. *Records Management Journal*, 30(2), 155-174.
- Jones, D. (2018). *3 façons d'utiliser l'intelligence artificielle pour redynamiser vos systèmes d'information en 2019*. Nuxeo. <https://www.nuxeo.com/fr/blog/ai-information-management>
- Lamouroux, M. et Ferchaud, B. (2006). Journée d'étude ADBS : L'impact du numérique sur l'évolution des modes de travail. *Documentaliste-Sciences de l'Information*, 43(3-4), 242-246.

- Lavigueur, É. (2020). L'essor des sciences de l'information au Québec : Évolution des métiers, innovation des pratiques et engagement des professionnels de l'information dans la société québécoise. *I2D - Information, données et documents*, 1(1), 7-14.
- Rolan, G., Humphries, G., Jeffrey, L., Samaras, E., Antsouпова, T. et Stuart, K. (2019). More human than human? Artificial intelligence in the archive. *Archives and Manuscripts*, 47(2), 179-203.
- Schallier, W. (2019). De Gand à Santiago : mon voyage en bibliothèques. *I2D - Information, données et documents*, 2(2), 51-55.
- Secrétariat du Conseil du trésor (SCT) (2004). *Standard du gouvernement du Québec sur les ressources informationnelles. Projet. Gestion documentaire* (Extraits du CCI version 1.2). Secrétariat du Conseil du trésor, Québec.
- Servais P. et Mirguet, F. (dir.). (2015). *Archivistes de 2030. Réflexions prospectives*. Louvain-la-Neuve/Paris, Academia-L'Harmattan.
- TechTarget (2018). *Traitement du langage naturel (TLN ou NLP)*. LeMagIT. <https://www.lemagit.fr/definition/Traitement-du-langage-naturel-TLN#:~:text=Le%20traitement%20du%20langage%20naturel,des%20technologies%20d'intelligence%20artificielle>
- Theimer, K. (2018). It's the end of archival profession as we know it, and I feel fine. Dans C. Brown (dir.), *Archival futures* (p. 1-18). Londres, Facet Publishing.
- Utard, J.-C. (2017). Bibliothécaire, documentaliste et archiviste : Convergence ou hybridation des métiers? Dans N. Marcerou-Ramel (dir.), *Les métiers des bibliothèques* (p. 157-166). Paris, Éditions du Cercle de la Librairie.
- Zwarich, N. et Maurel, D. (2020). La gouvernance de l'information et ses impacts sur la formation des professionnels de l'information au Québec. *I2D - Information, données et documents*, 1(1), 51-55.