

ARCHIVAGE DES DONNEES NUMERIQUES MASSIVES : CONTRIBUTION DE L'INTELLIGENCE ARTIFICIELLE ET DE LA BLOCKCHAIN

Reine Marie Ndéla MARONE

Laboratoire de Recherche en Sciences de l'Information et de la Communication (LARSIC)

EBAD, Université Cheikh Anta Diop de Dakar, Sénégal

reinemarie.marone@ucad.edu.sn

&

Moustapha MBENGUE

Laboratoire de Recherche en Sciences de l'Information et de la Communication (LARSIC)

EBAD, Université Cheikh Anta Diop de Dakar, Sénégal

moustapha.mbengue@ucad.edu.sn

&

Djibril DIAKHATÉ

Laboratoire de Recherche en Sciences de l'Information et de la Communication (LARSIC)

EBAD, Université Cheikh Anta Diop de Dakar, Sénégal

djibril.diakhate@ucad.edu.sn

Résumé : Le volume d'informations produites dans les entreprises ne cesse de croître à une vitesse très rapide du fait de la mutation de leur système de production, induite par l'utilisation des technologies de l'information et de la communication. Dans les structures publiques et privées, la difficulté d'accéder en temps réel à la documentation administrative est une réalité. Dans ce contexte, deux questions se posent. D'abord comment extraire de cette masse documentaire les informations utiles à l'activité ? L'autre interrogation se rapporte à la question de l'authenticité, mais aussi de la conservation sur le long terme des documents au format numérique, sachant que les technologies évoluent rapidement. Dans cet article, notre objectif est de montrer que l'intelligence artificielle (IA) et la blockchain peuvent apporter des éléments de réponse à ces problématiques. Notre travail s'inscrit dans le cadre théorique de l'intelligence artificielle et de la blockchain dans l'archivage des Big data et utilise la méthodologie de l'analyse qualitative. Les résultats de l'étude montrent que l'IA permet d'exploiter les données archivées de manière plus efficace et de redonner vie à un patrimoine informationnel sous-exploité, tandis que la blockchain peut aider à valider ou certifier les documents afin de garantir leur immuabilité et leur non-altération.

Mots clés : Intelligence Artificielle, Archives, Blockchain, Big Data.

Abstract: The volume of information produced in companies continues to grow at a very rapid rate due to the changes in their production systems brought about by the use of information and communication technologies. In public and private structures, the difficulty of accessing administrative documentation in real time is a reality. In this context, two questions arise. Firstly, how to extract from this mass of documentation the information that is useful for the activity. The other question relates to the issue of authenticity, but also to the long-term

preservation of documents in digital format, given that technologies are evolving rapidly. In this article, we try to see to what extent artificial intelligence (AI) and blockchain can provide answers to these problems. Our work is based on the theoretical framework of artificial intelligence and blockchain in the archiving of Big Data and uses the methodology of qualitative analysis. The results of the study show that AI can be used to exploit archived data more efficiently and to revive an underused information heritage, while blockchain can help validate or certify documents to guarantee their immutability and non-alteration.

Keywords: Artificial Intelligence, Archives, Blockchain, Big Data.

Introduction

Les Big data ou données de masse, désignent des « flux de données numériques variées, qui peuvent être de nature textuelle ou audiovisuelle et qui sont produites en masse par l'activité quotidienne des entreprises privées ou publiques, ou des organismes scientifiques » (Espinasse et Bellot, 2017). Elles peuvent être issues des instruments de mesure, des processus de travail collectif, des publications en ligne ou des collectes de l'activité des internautes. L'intérêt de l'exploitation des Big data, est de pouvoir tirer profit de ces données produites par les acteurs (entreprises, particuliers, scientifiques, institutions publiques) afin de permettre aux entreprises d'optimiser leur offre commerciale, leurs services, mais également développer la recherche et créer des emplois.

Les Big data touchent tous les domaines d'activité, et leur organisation, leur mise en valeur, est au cœur des problématiques actuelles de la recherche en science de l'information et de la communication, plus particulièrement chez les professionnels de l'information documentaire (archivistes, bibliothécaires, documentalistes, etc) (Ben Amor et Mkadmi, 2018). En effet, beaucoup d'informations numériques sont manipulées dans ces derniers : par exemple des documents scannés, des factures électroniques ou encore des documents de travail archivés dans des serveurs. Face à cette masse de données qui ne fait que croître, la principale difficulté qui se pose est de savoir comment extraire des informations utiles, c'est-à-dire comment exploiter les données dans une perspective de performance et de compétitivité. Par exemple : comment retrouver, en temps réel, les informations nécessaires parmi une multitude de mails, de documents, de fichiers multimédias, etc., surtout si le travail de classement et d'indexation préalable n'a pas été fait ?

À cela s'ajoute le problème de la conservation à long terme des documents au format numérique, dû à l'obsolescence rapide des technologies informatiques. Le potentiel besoin de migrer ces données à travers les technologies matérielles et

logicielles soulève des questions liées à l'authenticité des documents. Les méthodes de conservation à long terme peuvent éventuellement compromettre l'authenticité et l'intégrité des données. L'altération est en effet un risque encouru dans le passage d'un format à un autre puisqu'il est possible que des données aient été modifiées intentionnellement ou par erreur.

Quelle que soit l'intelligence mise autour du document, les questions immuables sur la validité, l'intégrité et la traçabilité de chaque document persisteront. Il est important de prouver leur provenance et leur intégrité au public de manière sûre et transparente.

Au cours des dernières années, les fournisseurs de technologies ont introduit de nouvelles technologies d'analyse et de traitement des Big Data telles en ayant recours à l'intelligence artificielle et à la blockchain qui permettent aux entreprises d'exploiter la valeur d'énormes volumes de données (Mougayar, 2016 ; Phillipon et Besson, 2020).

Il convient dès lors de s'interroger sur la contribution de l'intelligence artificielle et la blockchain dans l'archivage des big data.

Nous postulons que le développement des technologies d'intelligence artificielle et celles de la blockchain constituent une opportunité pour les services d'archives et les archivistes. Outil d'aide à l'archivage, l'IA peut permettre de mieux exploiter les données archivées en redonnant vie à un patrimoine informationnel sous-exploité grâce à des recherches basées non plus uniquement sur la sémantique, mais intégrant les données contextuelles en se plaçant du point de vue de la connaissance. La blockchain, quant à elle peut aider à valider ou certifier les documents afin de garantir leur immuabilité, leur non-altération.

Cependant, malgré le fait que ces différentes technologies sont une opportunité pour les services d'archives et les archivistes, elles ne peuvent pas remplacer ces derniers. Dans cet article, nous allons, par une démarche méthodologique qualitative, analyser la contribution de l'intelligence artificielle et de la blockchain dans l'archivage des mégadonnées. Ce qui va nous amener à discuter les limites possibles de ces technologies et du rôle que l'archiviste peut jouer pour aider à les combler. Auparavant, dans la première partie, nous reviendrons sur le cadre théorique et conceptuel de l'étude.

1. Cadre conceptuel et théorique

Notre étude s'inscrit dans le cadre théorique de l'intelligence artificielle et de la blockchain dans l'archivage des Big data. Il convient dès lors de faire une recension des publications dans ce domaine. Depuis quelques années, l'Intelligence artificielle (IA) fait l'objet d'une médiatisation et d'une attention particulière. De nombreux secteurs sont concernés par l'IA y compris l'archivistique et la bibliothéconomie. L'IA est le plus souvent perçue comme un ensemble de technologies pouvant produire de nombreux bénéfices, notamment en termes de performance (optimisation de processus internes, rapidité d'exécution de tâches, accroissement de la productivité, etc.) et peut faciliter le travail en permettant l'automatisation de certains traitements fastidieux ou répétitifs (Zouinar, 2020). Cela s'applique à de nombreux domaines d'activité tels que le traitement documentaire comme l'affirme un grand nombre de dirigeants d'entreprise, qui pensent à l'automatisation des processus documentaires. En effet selon Bruno TEXIER (2019 a). : « 72 % des entreprises s'attendent à ce que leur volume de données soit multiplié par trois (ou plus) dans les deux prochaines années. De ce fait, 71 % des dirigeants d'entreprise pensent que les processus documentaires pourraient être automatisés. »

Aussi, des entreprises comme Spigraph, Kofax et DS Informatique, considèrent que l'intelligence artificielle libérera les entreprises des traitements documentaires devenus incessants (Texier, 201b) et ont commencé à développer des solutions automatisées de traitement du courriel entrant. En plus d'aider à l'automatisation de certains processus documentaires, l'IA permet également d'accroître la valeur globale du patrimoine documentaire des entreprises, selon Alain Escaffre, directeur des produits au sein de Nuxeo (Texier, 2019c). Un avis partagé par le CR2PA (club des responsables de politiques et projets d'archivage), « l'IA peut également être un moyen de mieux exploiter les données archivées en redonnant vie à un patrimoine informationnel sous-exploité grâce à des recherches basées non plus uniquement sur la sémantique, mais intégrant les données contextuelles en se plaçant du point de vue de la connaissance ». Poursuivant sur la même idée le CR2PA soutient que l'IA va permettre une meilleure traçabilité et contribuer à une gestion plus efficiente de l'archivage » (CR2PA, 2019).

La blockchain, quant à elle, a récemment reçu une attention particulière en raison de sa fourniture de services de partage de données sécurisés avec traçabilité, immuabilité et non-répudiation.

Une blockchain est une base de données distribuée, qui maintient une liste croissante d'enregistrements (blocs), qui sont protégés contre toute révision ou altération non autorisée. Chaque bloc a un lien vers un bloc précédent. La technologie blockchain est censée établir l'authenticité des enregistrements. Elle génère également de nouveaux enregistrements « en chaîne, tels que les contrats intelligents (Marciano et al., 2018).

L'Université de Surrey utilise d'ailleurs les technologies blockchain associées à l'intelligence artificielle pour sécuriser les archives gouvernementales numériques. Les archives nationales du Royaume-Uni, de l'Australie et des aux États-Unis d'Amérique travaillent sur des projets similaires. Le professeur John Collomosse, qui dirige ce projet à l'Université de Surrey, a déclaré : « *Les archives à travers le monde amassent de vastes volumes de contenu numérique, et il est important qu'elles puissent prouver leur provenance et leur intégrité au public de manière sûre et transparente). En combinant les technologies de la blockchain et de l'intelligence artificielle, nous avons montré qu'il était possible de sauvegarder l'intégrité des données d'archives à l'ère numérique. La blockchain fournit essentiellement une empreinte digitale pour les archives, permettant de vérifier leur authenticité.* » (Bell et al., 2019).

2. Méthodologie de recherche

L'approche qualitative privilégiée dans cette étude est plus adaptée pour la réalisation d'un état de l'art, car permettant d'étudier les principales connaissances relatives à l'objet de recherche et d'en dresser un portrait complet. Le type d'analyse qualitative sur lequel notre choix s'est porté est : l'analyse thématique, qui nous a permis de faire la recension des écrits sur l'intelligence artificielle et la blockchain dans l'archivage des big data. En analyse qualitative, l'analyse thématique permet à partir d'un procédé de codification (ou classification) de faire ressortir les thèmes centraux (étape de thématisation) des références étudiées (corpus d'entretiens, de textes, etc.). Il s'agit donc de transposer un corpus donné en un certain nombre de thèmes représentatifs du contenu analysé, et ce, en rapport avec l'orientation de recherche. Le procédé de codification correspond à la segmentation du contenu d'un texte, d'un article, d'un entretien « en unités d'analyse (mots, phrases, thèmes...) et à leur intégration dans des thèmes sélectionnés en fonction de l'objet de recherche » (Thiétart, 2007, page 455).

Selon Pierre Paillé et Alex Muchielli (2008, p. 162), « *l'analyse thématique consiste à procéder systématiquement au repérage, au regroupement et subsidiairement, à l'examen discursif des thèmes abordés dans un corpus* ». Pour ces auteurs, elle « *une démarche de réduction des données à l'aide de thèmes par un travail systématique de synthèse des propos* » (exemple : la littérature scientifique, grise et professionnelle).

Ainsi, l'analyse thématique permet, de trier et fouiller le contenu d'un corpus, selon ses ressemblances « thématiques », de rassembler ces éléments dans des classes conceptuelles, d'étudier les relations existantes entre ces éléments et de donner une description compréhensive de l'objet d'étude » (Wanlin, 2007, p. 252). Cela nécessite alors de constituer au préalable, le corpus (c'est-à-dire un ensemble de textes, de documents, significatifs et représentatifs du sujet étudié, assemblé dans le but d'étudier son contenu, ses résultats). Ce corpus sélectionné permettra au chercheur de découvrir et d'appréhender son objet de recherche ainsi que de faire état précisément des connaissances et des méconnaissances actuelles du sujet. Avant de procéder à la recherche des documents, il est important de définir d'abord l'objectif principal de l'état de l'art ainsi que les indicateurs qui permettent de retenir les différents textes et documents composant le corpus. Dans notre cas, le but principal de la recension des écrits est de dresser un état des lieux des travaux sur l'intelligence artificielle et la blockchain dans l'archivage des big data. Cette recension doit permettre d'engager une discussion concernant les manquements de la littérature scientifique sur ce sujet. Après avoir défini l'objectif de la recension, il faut préciser les indicateurs d'inclusion et d'exclusion retenus pour intégrer ou non au corpus les documents à étudier. Ceux-ci permettent de préciser et limiter les contours du corpus de recherche. Dans notre travail de recension des écrits, nous avons choisi des indicateurs en tenant compte de la diversité des sources scientifiques, institutionnelles et professionnelles (voir figure 1). Notre objectif est de tenir compte des connaissances empiriques qui peuvent être diffusées à travers les données des littératures scientifique, professionnelle et grise.

<i>Type de Source</i>	Articles/Ouvrages/Mémoires/Rapports scientifiques, professionnels, institutionnels
<i>Objet des travaux</i>	Les données doivent faire référence à l'utilisation de l'intelligence artificielle ou celle de la blockchain dans l'archivage des big data
<i>Indicateur de fiabilité</i>	Importance de la fiabilité de la source, de sa diffusion et de sa visibilité dans le champ scientifique et professionnel
<i>Indicateur d'accès</i>	Disponibilité de la donnée bibliographique, accessibilité financière

Tableau 1 – Indicateurs choisis pour la composition du corpus

Nous avons effectué notre recherche de données bibliographiques grâce à une recherche basée sur des mots-clés relatifs à notre thème d'étude sur les bases de données scientifiques (CAIRN, Erudit, Persée), et aussi sur les réseaux sociaux scientifiques (ResearchGate et Académia)

Les mots-clés qui ont été retenus pour la recherche sont :

- Intelligence Artificielle et Archivage,
- Blockchain et Archivage,
- Intelligence Artificielle et Archivage Big Data,
- Blockchain et Archivage Big Data,

Les résultats des recherches menées sur ces différentes plateformes nous ont permis de prendre connaissance des références bibliographiques intéressantes par rapport à notre sujet de recherche.

Sur certaines plateformes, on a pu accéder directement aux fichiers (en format PDF ou traitement de texte) en les téléchargeant. Pour chaque document accédé, nous avons effectué une préanalyse de sa source en nous appuyant sur les indicateurs définis en amont. Nous avons ensuite évalué le document au regard de son intérêt scientifique pour la réalisation de l'état de l'art. Si un document correspond aux indicateurs fournis en amont, celle-ci est alors intégrée au corpus de document dans la démarche de recension des écrits. À l'issue de cette première étape de création du corpus par le recueil des données, seuls quelques documents ont été retenus (après l'étape de la préanalyse). En effet, au cours de nos recherches, nous avons constaté que le sujet : Intelligence Artificielle ou Blockchain dans l'archivage des données n'a pas été beaucoup abordé. Pour organiser la recherche des documents, il était essentiel de pouvoir classifier les différentes données afin de faciliter leur exploitation ultérieure. Pour cela, nous avons utilisé le logiciel de gestion bibliographique Zotero (Puckett, 2011). Dans la réalisation de l'état de l'art, Zotero nous a facilité le recueil des références bibliographiques et le classement de celles-ci. Une fois le corpus constitué, il a été possible de procéder à l'opération suivante : le codage des données bibliographiques. La lecture des documents de notre corpus et les opérations de codage du contenu ont permis de mettre en exergue principalement 3 thèmes. Les thèmes identifiés ont abordé le rôle de l'intelligence artificielle ou de la blockchain dans l'archivage des big data.

Ces 3 thèmes retenus sont :

1. la classification automatique de documents à archiver
2. la recherche d'informations améliorée par l'IA
3. la validation ou certification des documents d'archives avec la blockchain

À partir de là, nous sommes alors passés aux étapes suivantes d'interprétation, de discussion des résultats et de la rédaction de l'état de l'art.

3. L'IA et la Blockchain dans l'archivage

Les technologies de l'intelligence artificielle et celles de la blockchain peuvent grandement participer à révolutionner le monde, mais leurs applications restent très techniques et pas très connues du grand public, même si l'intelligence artificielle est de plus en plus connue (Salmeron, 2019). La blockchain et l'intelligence artificielle sont des technologies différentes, mais pouvant être combinées pour disposer d'un vaste panel d'applications utiles à l'archivage.

3.1. *L'intelligence artificielle au service de l'archivage*

L'intelligence artificielle est une discipline scientifique qui consiste à mettre en œuvre un certain nombre de techniques afin de permettre aux machines d'imiter une forme d'intelligence humaine. Avec des méthodes d'apprentissage, l'intelligence artificielle analyse des données afin d'organiser les informations et apprendre à résoudre des problèmes. Dans les entreprises, beaucoup d'informations sont manipulées à des fins d'archivage, de stockage et de diffusion, c'est le cas des documents scannés, des factures électroniques ou encore tous les documents de travail stockés dans des serveurs. Cependant avec cette masse de données croissante, trouver les bonnes informations en temps voulu s'avère de plus en plus difficile. L'intégration de l'intelligence artificielle au sein des systèmes d'information des entreprises peut apporter des éléments de réponse à cette problématique. Cette technologie permet aux ordinateurs de comprendre, d'interpréter et de traiter une multitude d'informations de façon autonome, mais également de faciliter la recherche dans cette masse de données. Grâce à l'IA l'ordinateur est aujourd'hui capable de catégoriser automatiquement des documents scannés sans pour autant qu'une personne ne soit obligée de les ouvrir un par un, de les lire et de les affecter manuellement à une catégorie (indexation et classification), ce qui permettra par la suite de pouvoir les

retrouver plus facilement. Cette classification automatique des documents, peut se faire sans pour autant définir de modèles, de mots-clés, ou d'index.

L'IA peut extraire la bonne information et donner du sens à une multitude de documents, quels que soient leur taille, leur format, leur langue ou les symboles utilisés. Un des autres avantages de l'IA est qu'elle élargit les possibilités de la recherche d'informations contrairement aux moteurs de recherche habituels. En effet, dans les moteurs de recherche classiques la recherche d'informations est généralement basée sur la saisie de mots-clés.

Cependant, le plus souvent celle-ci échoue si elle doit prendre en compte les photos, les plans, les illustrations, les fichiers vidéo ou audio. De plus, s'il y a des relations entre les documents il est nécessaire de les définir manuellement au moment de leur stockage, sinon lors de la recherche ces liens ne pourront pas être détectés. Par exemple la recherche d'information sur le premier président du Sénégal, Léopold Sédar Senghor, risque de ne retourner qu'un petit nombre des documents réellement disponibles. Cependant, si l'intelligence artificielle est utilisée, le système reconnaîtra, en plus des images sur le Président Senghor, ses réseaux de relations, par exemple sa connexion avec Mamadou DIA qui fut son Premier ministre.

La reconnaissance des images se fait grâce à la vision artificielle, qui a fait de grands progrès dernièrement. Cette technologie permet à la machine de déterminer précisément le contenu d'une image pour ensuite la classer automatiquement selon l'objet, la couleur ou le visage repéré. Ainsi, celle-ci devient capable de reconnaître, parmi des milliards d'images, une personne, un animal et bien d'autres catégories d'éléments. L'IA rend même possible la recherche sur la vidéo et le son. Ainsi, en combinant des technologies telles que la reconnaissance d'image, les logiciels parole-texte (speech-to-text en anglais), le traitement naturel du langage (natural language processing en anglais), l'IA peut automatiser la création de métadonnées sur des contenus textes ou multimédias, afin d'améliorer leur archivage et favoriser leur découverte. Récemment des projets ont été engagés dans le but de combiner intelligence artificielle et archivage de documents. C'est le cas du projet de recherche HIMANIS (HISTORICAL MANUSCRIPT INDEXING FOR USER-CONTROLLED SEARCH) mené par l'IRHT en 2015, en partenariat avec les archives nationales de France, qui a pour objet d'étude les registres de la chancellerie royale des XIVe et XVe siècles (Stutzmann et al.,

2017). L'enjeu technologique majeur était de procéder à une reconnaissance par ordinateur des écritures médiévales qui soient suffisamment fiables pour offrir à la communauté un moteur de recherche en plein texte dans les registres, dont une partie n'est pas encore couverte par les inventaires ou les index. Une interface de consultation des données, conjuguant un accès aux images, aux transcriptions et aux analyses archivistiques, a été également développée.

3.2. *La blockchain pour certifier les documents d'archives*

L'obsolescence rapide des technologies informatiques crée des difficultés pour ceux qui se soucient de la conservation à long terme des documents au format numérique. Le potentiel besoin de migrer ces données à travers les technologies matérielles et logicielles soulève des questions liées à l'authenticité des documents. Il est alors nécessaire de trouver des moyens qui permettront de garantir que les données n'ont pas été modifiées intentionnellement ou par erreur, et également de veiller à ce que les méthodes de conservation à long terme ne compromettent pas l'authenticité des données.

En plus de la problématique de l'authenticité, la question de la traçabilité des documents persiste également. En effet, quelle que soit l'intelligence mise autour du document, les questions suivantes demeurent : validé et certifié par qui ? Transmis à qui ? Déposé avec quelle preuve ?

Les technologies de type blockchain peuvent aider à fournir des éléments de réponse à ces problématiques.

La blockchain a récemment fait l'objet d'une attention particulière en raison de sa fourniture de services de partage de données sécurisés avec traçabilité, immuabilité et non-répudiation (Marciano et al., 2018).

Une blockchain est une base de données distribuée qui maintient une liste croissante d'enregistrements (blocs) protégés contre toute révision ou falsification non autorisée (Marciano et al., 2018). Chaque bloc a un lien vers un bloc précédent. Cette chaîne d'informations devient une base de données mondiale accessible à tous les utilisateurs. La blockchain est l'élément central de la monnaie numérique bitcoin et il sert de grand livre pour les transactions.

La blockchain fonctionne sans aucun organisme de contrôle. L'authenticité de l'information est garantie par les utilisateurs de la blockchain qui la valide grâce à la

cryptographie. À chaque transaction faite par un utilisateur en implémentant une information dans la blockchain, celle-ci est regroupée avec les autres informations connexes au sein d'un même bloc. Cette information est alors vérifiée et approuvée par les autres membres de la blockchain en utilisant des techniques cryptographiques. Ces utilisateurs, qui vérifient et valident la cohérence avec les autres informations du bloc, sont rémunérés. Lorsque le bloc est validé, il est horodaté et ajouté dans la blockchain. Dès lors, l'information devient visible et accessible à tous, mais ne pourra cependant plus être modifiée. Toute rectification d'information doit en effet s'effectuer par une nouvelle transaction.

Dans les blockchains modernes, il existe des programmes appelés « Smart Contract » ou contrats intelligents, qui permettent de contrôler une partie de la blockchain en organisant les relations entre les différentes parties et en établissant des actions telles que l'automatisation de certaines activités de gestion comme la (re) demande de document en cas d'expiration de validité. L'objectif est d'améliorer l'existant dans le but de correspondre aux nouvelles attentes du marché comme la transparence de l'information.

On peut donc constater que la technologie Blockchain est une nouvelle manière d'administrer la preuve et représente un réseau ouvert et décentralisé reposant sur un protocole de cryptographie (Marciano et al., 2018). Grâce à la blockchain, des organisations autonomes distribuées peuvent réaliser des transactions infalsifiables. Elle représente un moyen pour établir l'authenticité des documents (Marciano et al., 2018). La technologie blockchain offre plusieurs avantages. Elle permet le partage d'un même document entre tous les acteurs concernés, et cela, sans dupliquer le document et en assurant la maîtrise de sa visibilité. Grâce à la blockchain, il est possible de mettre en place un registre unique contenant tous les documents que nous avons l'habitude de fournir aux administrations, aux employeurs et aux entreprises commerciales avec lesquels nous sommes en relation. Les entreprises peuvent également utiliser la blockchain pour créer des registres spécialisés (données identité, données santé, données patrimoniales, données judiciaires...) afin de pouvoir diminuer la quantité de documents qu'ils gèrent pour leurs interactions B2B (Business to business) et B2C (business to consumer). De plus la maîtrise de ces documents sera octroyée à l'écosystème et pas à un acteur unique. La technologie blockchain peut permettre ainsi

de gérer la traçabilité des documents à archiver, les horodater et conserver l'ensemble des accès qui ont été effectués sur ces derniers. Elle fournit des mécanismes de validation/certification des documents signés (« Proof of Existence ») et peut aider à garantir l'immutabilité, la non-altération des documents via leur empreinte unique conservée dans un block.

Les Archives nationales britanniques (TNA) ont d'ailleurs lancé, en juin 2018, un projet de recherche baptisé ARCHANGEL, visant à utiliser la technologie blockchain pour archiver leurs documents et vérifier leur authenticité. ARCHANGEL a trois objectifs principaux : vérifier la provenance et s'assurer de l'intégrité sur le long terme des documents archivés, pérenniser la confiance du public dans le système des archives publiques, et créer une base de données entretenue en collaboration avec d'autres systèmes d'archives à travers le monde (Collomosse et al., 2018). La blockchain d'ARCHANGEL fonctionne comme une base de données maintenue par plusieurs archives ; tout le monde peut vérifier et ajouter des enregistrements, mais personne ne peut les modifier. Comme aucune donnée ne peut être modifiée, l'intégrité du dossier historique demeure intacte.

John Sheridan, directeur numérique des Archives nationales du Royaume-Uni, a déclaré : « *En explorant la technologie de la blockchain avec certaines des principales archives mondiales, le projet ARCHANGEL a montré, pour de vrai, comment les archives pourraient combiner leurs forces pour protéger et garantir des preuves numériques vitales pour l'avenir. ARCHANGEL est un partenariat exceptionnel qui a fourni des recherches révolutionnaires sur les aspects pratiques de l'utilisation de la blockchain pour garantir la confiance dans les archives numériques à grande échelle* » (Collomosse et al., 2018).

On peut donc dire que la blockchain pourrait être d'une grande utilité pour les services d'archives, elle permettrait par exemple d'aider à valider ou certifier les documents afin de garantir leur immutabilité, leur non-altération. Elle peut aussi permettre de garantir la traçabilité des documents à archiver, l'horodatage et la conservation de l'ensemble des accès qui ont été effectués sur ces derniers.

4. Discussion

Comme nous l'avons vu dans les parties précédentes, l'intelligence artificielle (IA) offre un avantage concurrentiel par rapport aux moteurs de recherche habituels, car elle élargit les possibilités de la recherche d'informations. De plus, les systèmes basés

sur l'IA apprennent de manière autonome et se perfectionnent constamment dans l'accomplissement de leurs tâches.

La blockchain de son côté est utilisée afin de valider et certifier les données archivées. Cependant, malgré le fait que ces différentes technologies soient une opportunité pour les services d'archives et les archivistes, elles ne peuvent pas remplacer ces derniers.

En effet si nous prenons le cas de la technologie blockchain, cette dernière permet de chaîner, partager, horodater et sécuriser les documents en préservant leur intégrité, leur fiabilité, leur authenticité et leur lisibilité. Mais peut-on vraiment faire confiance à la blockchain ?

Pour que la technologie blockchain puisse être considérée comme fiable, il est fondamental de garantir sa capacité à vérifier et auditer des enregistrements fiables. En d'autres termes, il est nécessaire de développer les critères par rapport auxquels la fiabilité de la blockchain peut être évaluée (Marciano et al., 2018). Pour cela, il faudrait s'inspirer des théories, principes et pratiques des sciences archivistiques et les normes internationales qui en sont issues. En effet, la théorie archivistique aborde les concepts clés nécessaires pour produire et conserver des documents fiables : exactitude, fiabilité et l'authenticité liée au concept de provenance (Marciano et al., 2018). Le développement et la mise en œuvre de la blockchain, qui est essentiellement une technologie de tenue de documents, sont peu susceptibles de réaliser les innovations et les perturbations tant médiatisées par ses partisans, tant qu'ils ne prennent pas en compte les connaissances archivistiques. De plus, ne pas prendre en compte ces connaissances archivistiques peut engendrer des conséquences graves et inattendues pour la société, telle que la perte de documents créés et potentiellement stockés sur la chaîne (comme les contrats intelligents) ou même une incapacité à établir l'authenticité des documents revendiquant des droits tels que le droit d'auteur et propriété foncière (Lemieux, 2016 a ; Lemieux, 2016b ; Lemieux, 2016 c).

Conclusion

Les entreprises modernes sont inondées de données dont le volume s'accroît à un rythme qui peut être très rapide. Il en va de même pour les archives d'une compagnie, qui peut très vite ne plus être capable de les questionner efficacement. Dans ce contexte, l'un des principaux problèmes est de savoir comment extraire des informations utiles, c'est-à-dire comment exploiter les données. Un autre défi majeur

est comment garantir leur authenticité, mais aussi comment assurer la conservation à long terme des documents au format numérique, sachant que les technologies peuvent rapidement devenir obsolètes. L'intelligence artificielle (IA) et la blockchain peuvent apporter des éléments de réponse à ces problématiques.

L'IA permet en effet d'exploiter les données archivées de manière plus efficace et de redonner vie à un patrimoine informationnel sous-exploité. Cette technologie permet aux ordinateurs, de comprendre, d'interpréter et de traiter une grande quantité d'informations de façon autonome et facilite la recherche dans ce tas d'information. Ainsi, un ordinateur pourra catégoriser automatiquement des documents scannés sans pour autant qu'un être humain ne soit obligé de les ouvrir un par un, de les lire et de leur attribuer manuellement une catégorie, ce qui permettra par la suite de pouvoir les retrouver plus facilement. La blockchain quant à elle, peut aider à valider ou certifier les documents afin de garantir leur immuabilité, leur non-altération. Elle peut aussi permettre de garantir la traçabilité des documents à archiver, l'horodatage et la conservation de l'ensemble des accès qui ont été effectués sur ces derniers. Toutefois, le développement et la mise en œuvre de la blockchain, sont peu susceptibles de réaliser les innovations et les perturbations tant médiatisées, tant qu'ils ne prennent pas en compte les connaissances archivistiques. En effet, la technologie à elle seule, ne peut pas tout faire, il est nécessaire de l'associer à l'expertise de l'archiviste, ses méthodes ainsi que son approche. Ne pas en tenir compte pourrait entraîner des conséquences graves et inattendues telles que la perte de documents créés et potentiellement stockés sur la chaîne (tels que les contrats intelligents) ou même une incapacité à établir l'authenticité des documents revendiquant des droits tels que le droit d'auteur et propriété foncière (Lemieux, 2016 a ; Lemieux, 2016 b ; Lemieux, 2016 c).

Dès lors on peut se poser la question suivante : n'est-il pas nécessaire que les archivistes soient mieux impliqués dans la conception des systèmes documentaires intégrant la blockchain ou l'intelligence artificielle ?

Références bibliographiques

- Ben Amor, Fatma, et Abderrazak Mkadmi. 2018. « Les Archives à l'Ère des Big Data: Les Enjeux de l'Archivage des Données Numériques Massives ». Dans *Proceedings of the 1st International Conference on Digital Tools & Uses Congress.*, 1-6.
- Collomosse, John, Tu Bui, Alan Brown, John Sheridan, Alex Green, Mark Bell, Jamie Fawcett, Jez Higgins, et Olivier Thereaux. 2018. « ARCHANGEL: Trusted archives of digital public documents ». Dans *Proceedings of the ACM Symposium on Document Engineering 2018.*, 1-4.
- CR2PA. 2019. « L'intelligence artificielle est-elle une menace ou une chance pour les professionnels en charge de l'archivage des données des entreprises et organisations ? » CR2PA. 15 juillet. <http://blog.cr2pa.fr/2019/07/lintelligence->

[artificielle-est-elle-une-menace-ou-une-chance-pour-les-professionnels-en-charge-de-larchivage-des-donnees-des-entreprises-et-organisations/](#).

- Espinasse, Bernard, et Patrice Bellot. 2017. « Introduction au Big Data - Opportunités, stockage et analyse des mégadonnées ». *Documents numériques Gestion de contenu*. <https://doi.org/10.51257/a-v1-h6040>.
- Green, Alex, Mark Bell, John Sheridan, John P. Collomosse, Tu Bui, Alan Brown, Jamie Fawcett, Olivier Thereaux, et Jeni Tennison. 2018. « Using blockchain to engender trust in public digital archives. » Dans *IPres 2018, International Conference on Digital Preservation*. Boston (USA). <https://files.osf.io/v1/resources/kefj8/providers/osfstorage/5bd739fd133ddd0019342281?action=download&direct&version=1>.
- Mougayar, William. 2016. *The business blockchain: promise, practice, and application of the next Internet technology*. Hoboken (New Jersey) : John Wiley & Sons.
- Paillé, Pierre, et Alex Mucchielli. 2012. *L'analyse qualitative en sciences humaines et sociales*. Armand Colin.
- Pierrot, Raluca. 2020. « Entre intelligence artificielle et science ouverte: quelles évolutions du métier de bibliothécaire? » *Arabesques*, n° 96 (janvier) : 21. <https://doi.org/10.35562/arabesques.1487>.
- Puckett, Jason. 2011. *Zotero: A guide for librarians, researchers, and educators*. Assoc of Cllge & Rsrch Libr.
- Ramjaun, Rana. s. d. « Recherche Vocale : La Tendence Du Futur ». Myconnecting. Consulté le 19 novembre 2021. <https://www.myconnecting.fr/articles/recherche-vocale/>.
- Salemeron, Léa. 2019. « L'Intelligence artificielle et la blockchain peuvent-elles se combiner? » *MBA MCI*. 5 septembre. <https://mbamci.com/l-intelligence-artificielle-et-la-blockchain-se-combinent/>.
- Stutzmann, Dominique, Jean-François Moufflet, et Sébastien Hamel. 2017. « La recherche en plein texte dans les sources manuscrites médiévales: enjeux et perspectives du projet HIMANIS pour l'édition électronique ». *Médiévales. Langues, Textes, Histoire* 73 (73) : 67-96.
- Texier, Bruno. 2016. « Quand le document se dope à l'intelligence artificielle ». *Archimag*, 2016. <https://www.archimag.com/veille-documentation/2016/04/28/document-dope-intelligence-artificielle>.
- Texier, Bruno. 2019a. « ECM : l'intelligence artificielle enrichit la valeur des données stockées ». *Archimag*, 2019. <https://www.archimag.com/demat-cloud/2019/08/21/gestion-contenu-intelligence-artificielle-enrichit-valeur-donnees-stockees>.

- Texier, Bruno. 2019b. « Quand l'intelligence artificielle s'invite dans le traitement documentaire ». Archimag. 2019. <https://www.archimag.com/demat-cloud/2019/10/03/quand-intelligence-artificielle-invite-traitement-documentaire>.
- Thiétart, Raymond-Alain. 2007. *Méthodes de recherche en management Ed. 3*. 3^e éd. Paris, France : Dunod. <https://www.scholarvox.com/catalog/book/10232253>.
- Wanlin, Philippe. 2007. « L'analyse de contenu comme méthode d'analyse qualitative d'entretiens: une comparaison entre les traitements manuels et l'utilisation de logiciels ». *Recherches qualitatives* 3 (3) : 243-72.