

**CONCOURS EXTERNE, INTERNE ET TROISIÈME CONCOURS
D'ASSISTANT TERRITORIAL DE CONSERVATION DU PATRIMOINE
ET DES BIBLIOTHÈQUES**

SESSION 2025

ÉPREUVE DE NOTE

ÉPREUVE D'ADMISSIBILITÉ :

Rédaction d'une note à l'aide des éléments d'un dossier portant sur la spécialité choisie par le candidat au moment de l'inscription.

Durée : 3 heures
Coefficient : 3

SPÉCIALITÉ : ARCHIVES

À LIRE ATTENTIVEMENT AVANT DE TRAITER LE SUJET :

- ♦ Vous ne devez faire apparaître aucun signe distinctif dans votre copie, ni votre nom ou un nom fictif, ni initiales, ni votre numéro de convocation, ni le nom de votre collectivité employeur, de la commune où vous résidez ou du lieu de la salle d'examen où vous composez, ni nom de collectivité fictif non indiqué dans le sujet, ni signature ou paraphe.
- ♦ Sauf consignes particulières figurant dans le sujet, vous devez impérativement utiliser une seule et même couleur non effaçable pour écrire et/ou souligner. Seule l'encre noire ou l'encre bleue est autorisée. L'utilisation de plus d'une couleur, d'une couleur non autorisée, d'un surligneur pourra être considérée comme un signe distinctif.
- ♦ Le non-respect des règles ci-dessus peut entraîner l'annulation de la copie par le jury.
- ♦ Les feuilles de brouillon ne seront en aucun cas prises en compte.

Ce sujet comprend 27 pages.

**Il appartient au candidat de vérifier que le document comprend
le nombre de pages indiqué.**

S'il est incomplet, en avertir le surveillant.

Vous êtes assistant territorial de conservation du patrimoine au sein du service des archives de Cultureville (50 000 habitants).

L'élue déléguée à la Culture souhaiterait mesurer les impacts que pourrait avoir le développement de l'intelligence artificielle sur les différentes politiques publiques en matière culturelle.

Dans cette perspective, le Directeur de la culture vous demande de rédiger à son attention, exclusivement à l'aide des documents joints, une note sur le rôle de l'intelligence artificielle (IA) en matière d'archives.

Liste des documents :

Document 1 : « L'intelligence artificielle a sa place dans le monde des archives » - *archimag.com* - 20 février 2023 - 3 pages

Document 2 : « Lettres en lumière. L'intelligence artificielle au service du patrimoine écrit » (extraits) - *Archives départementales de Côte d'Or* - 2022 - 7 pages

Document 3 : « L'IA, c'est quoi ? » (extraits) - *Conseil de l'Europe - coe.int* - Consulté le 10 novembre 2023 - 3 pages

Document 4 : « L'intelligence artificielle pour gouverner les archives nationales américaines » - *actualité.com* - 20 avril 2021 - 1 page

Document 5 : « L'intelligence artificielle au service du traitement des archives » - *Ar(abes)ques* n°107 - Novembre-décembre 2022 - 2 pages

Document 6 : « Archivistique à l'ère de l'IA: Opportunités, défis et besoin d'utilisation responsable » (extraits) - Basma Makhoul Shabou - *Arbido.ch* - 2023 - 3 pages

Document 7 : « IA : "On va avoir de plus en plus de mal à constituer des archives authentiques" » (extraits) - *usbeketrica.com* - 26 septembre 2023 - 2 pages

Document 8 : « Recommandation sur l'éthique de l'intelligence artificielle » (extraits) - *UNESCO* - Novembre 2021 - 1 page

Document 9 : « L'intelligence artificielle est-elle une menace ou une chance pour les professionnels en charge de l'archivage des données des entreprises et organisations ? » - *blog.cr2pa* - 15 juillet 2019 - 1 page

Document 10 : « Éduquez l'intelligence artificielle avec les Archives de Belfort » - *rfgenealogie.com* - 6 mars 2023 - 2 pages

Documents reproduits avec l'autorisation du C.F.C.

Certains documents peuvent comporter des renvois à des notes ou à des documents non fournis car non indispensables à la compréhension du sujet.

"L'intelligence artificielle a sa place dans le monde des archives"

Le 20/02/2023 (Mis à jour le 09/11/2023) – *archimag.com* - Bruno Texier

L'intelligence artificielle a-t-elle quelque chose à apporter aux archives ? Françoise Banat-Berger, Emmanuelle Bermès et Stéphane Pouyllau partagent leurs points de vue sur les évolutions dans le domaine de l'archivage électronique et physique.

Le point de vue de Françoise Banat-Berger

Françoise Banat-Berger est directrice du service interministériel des Archives de France (Siaf).

Quels sont les prochains défis technologiques qui attendent les archivistes ?

Les archivistes sont aujourd'hui confrontés à deux défis majeurs. Le premier est la prise en charge d'archives numériques de plus en plus nombreuses (fichiers bureautiques, messageries) comme en témoignent notamment les derniers versements des cabinets ministériels, de productions audiovisuelles massives, ou également d'extraction de données à partir d'applications métiers ou de gestions électroniques de documents.

Les systèmes d'archivage électronique à même de prendre en charge ces données existent : le programme interministériel d'archivage électronique Vitam et la mise en production de l'offre de service Vitam accessible en service, que pilote aujourd'hui le ministère de la Culture, ainsi que d'autres solutions offertes par le secteur privé ou de l'économie sociale et solidaire.

La généralisation de leur déploiement est le premier enjeu. Le second consiste à « passer à l'échelle », c'est-à-dire à acquérir la capacité de verser en masse et régulièrement ces données. Un autre grand défi concerne l'accès aux archives à distance, sachant que, selon les données d'activité des services d'archives en 2021, on compte aujourd'hui un usager en salle de lecture pour 330 internautes. Les enjeux sont donc énormes, qu'il s'agisse de la visibilité des ressources archivistiques sur le web, qui repose pour nous sur l'énorme potentiel du portail FranceArchives, ou de la mise en place de solutions d'accès à distance sécurisé pour les ressources non encore diffusables sur internet.

L'intelligence artificielle (IA) a-t-elle sa place dans la fonction archive ?

L'IA a évidemment sa place dans le monde des archives, notamment autour de la reconnaissance automatique des caractères manuscrits. À cet égard, les Archives nationales ont été pionnières, avec une expérimentation sur les archives du Trésor des chartes, des inventaires du 18e siècle, ou enfin des répertoires de notaires.

Les Archives départementales lancent également des expérimentations, à l'image de celle menée par les Archives départementales de la Côte-d'Or.

Nous avons également lancé, dans le cadre d'un projet de recherche porté par l'Ined et la société Teklia, une vaste opération similaire sur l'ensemble des recensements de population entre 1836 et 1936 qui permettra d'extraire près de 500 millions de données nominatives.

Le point de vue d'Emmanuelle Bermès

Emmanuelle Bermès est responsable pédagogique du master « Technologies numériques appliquées à l'histoire », au sein de l'École nationale des chartes.

Quelle est la place de l'intelligence artificielle dans la fonction archive ?



L'IA a de nombreuses applications potentielles pour les professionnels de l'information : de la fouille d'images à la reconnaissance d'entités nommées, en passant par la prédiction de politiques de conservation. Dans les archives, le cas d'usage le plus prégnant à l'heure actuelle porte sur la reconnaissance de l'écriture manuscrite ou HTR (pour handwritten text recognition).

L'IA permet d'entraîner un outil à analyser la mise en page d'un document et reconnaître différents types d'écritures, ce qui permet d'extraire du texte semi-structuré à partir de documents anciens numérisés en mode image.

Vous êtes investie dans le programme AI4LAM (Artificial Intelligence for Libraries, Archives and Museums). De quoi s'agit-il ?

AI4LAM est une communauté internationale de professionnels des bibliothèques, archives et musées qui s'intéressent à l'utilisation de l'IA dans ces domaines. Des événements, formations et groupes de travail sont régulièrement proposés. Il existe un chapitre (ou sous-communauté) francophone que chacun peut rejoindre.

Quelle est la place des technologies numériques dans l'enseignement au sein de l'École nationale des chartes ?

La spécialité de l'École des chartes est l'analyse des documents de tous types et de toutes périodes, du Moyen-âge à nos jours : prendre en compte l'évolution des fonds et des pratiques documentaires est au cœur de ses missions.

Le master « Technologies numériques appliquées à l'histoire », qui existe depuis plus de 15 ans, forme de jeunes professionnels spécialistes du traitement de la donnée et du document numérique. Les méthodes des humanités numériques sont également enseignées dans la formation des archivistes paléographes et dans un master dédié, porté par l'École au sein de PSL.

Ces compétences très pointues sont notamment investies dans le Consortium pour la reconnaissance d'écriture manuscrite des matériaux anciens (Cremma), un consortium spécialisé dans le déploiement de l'HTR, dont l'École des chartes est l'un des membres fondateurs.

Le point de vue de Stéphane Pouyllau

Stéphane Pouyllau est ingénieur de recherche hors classe CNRS, co-fondateur d'Huma-Num, Isidore et Nakala, et responsable du HN Lab.

Vous avez été à l'origine de MédiHAL (archive ouverte pour photographies et images scientifiques). De quoi s'agit-il ?



MédiHAL, créée en 2009 avec Daniel Charnay, Laurent Capelli, du CCSD, et Shadia Kilouchi (CNRS) est une archive ouverte qui permet de déposer des données visuelles et sonores (images fixes, vidéos et sons). La particularité de MédiHAL est qu'il s'agit d'un entrepôt de données qui s'adresse à toutes les disciplines des sciences. Pour les communautés des sciences humaines et sociales, avoir une archive ouverte pour sauvegarder, documenter et valoriser les collections iconographiques des équipes de recherche était un besoin de plus en plus important.

Quels sont les défis à relever dans le domaine spécifique des archives scientifiques ?

Il y a un double aspect : la conservation des archives scientifiques, ce qui renvoie aux méthodes et techniques en archivistique et en même temps l'exploitation scientifique des contenus de ces archives. Si le signalement des fonds dans le portail Calames est important, il faut aussi permettre aux programmes de recherche de puiser de l'information dans les contenus de ces fonds.

Après la numérisation des fonds, il s'agit d'en permettre l'exploitation à grande échelle. Cela passe par l'utilisation de l'IA par exemple, avec les outils de HTR qui se développent fortement aujourd'hui dans les équipes de recherche en SHS.

L'intelligence artificielle a-t-elle quelque chose à apporter aux archivistes ?

Absolument. Outre l'exploitation des fonds numérisés, l'IA peut contribuer à la fabrication d'instruments de recherche pouvant offrir plusieurs types d'accès pour faire de la recherche. L'IA peut aider à passer d'instruments statiques (par exemple des collections d'instruments en EAD-XML) à la création de moteurs de recherche dynamiques sur ces instruments de recherche.

Au-delà, l'usage de l'IA permet de développer des outils de traitement de l'information à très grande échelle, et donc pour les fonds d'archives. Par exemple, avec l'aide de la Scop Dataactivist, nous avons développé des modèles d'IA pour le traitement des archives des réponses au Grand Débat de 2019.

Cela a permis d'offrir une vision des toponymes du Grand Débat ou encore d'aborder des questions de spatialisation et de localisation des références textuelles, etc. L'IA peut fournir des boîtes à outils intéressantes pour construire de nouvelles interactions entre archives et programmes de recherche.

Lettres en lumière. L'intelligence artificielle au service du patrimoine écrit (extraits) - Archives départementales de Côte d'Or

Chers contributeurs



Depuis 25 ans, les Archives en France numérisent le patrimoine et le mettent en ligne massivement ; l'objectif est désormais d'océriser ces images numériques (c'est-à-dire de les transcrire automatiquement), et les rendre ainsi accessibles en texte intégral, traduites si nécessaire en français moderne et enrichies d'une indexation performante.

Le projet porté par les Archives départementales de Côte-d'Or et l'Université de Bourgogne tend donc à mettre à la portée de tous le patrimoine écrit français, accessible aujourd'hui au mieux sous forme d'images en ligne, mais, le plus souvent, sous forme d'originaux consultables exclusivement dans les salles de lecture.

Il s'agit donc de lever les obstacles du temps (horaires d'ouverture), de l'espace (lieux de consultation) et du savoir-faire paléographique (difficultés de lecture) pour permettre l'accès à tous de documents historiques de grande valeur.

Si vous souhaitez nous aider à rendre accessible au plus grand nombre le très riche patrimoine écrit de France, à faire rayonner les archives de la Bourgogne en offrant aux chercheurs du monde entier un outil de travail unique, nous serons heureux de vous compter parmi nos partenaires.

Édouard Bouyé

Directeur des Archives
départementales
de la Côte-d'Or

Annie Vinter

Professeur à l'Université
de Bourgogne
Directrice de thèse

Les registres des États de Bourgogne : des documents à fort intérêt historique et paléographique.

Parmi les écritures anciennes, certaines sont très complexes, même pour des paléographes avertis : c'est le cas des écritures fin XVI^e-début XVII^e siècle. Au XVIII^e siècle, en revanche, un œil un peu exercé de 2021 lit la plupart des documents. L'écriture de cette période semble donc un bon point de départ pour nourrir un système d'océrisation, qui pourrait ensuite s'appliquer aux millions de pages du XVIII^e siècle conservées en France. On pourrait ensuite étendre le processus en amont et en aval, vers les XVII^e et XIX^e siècles.

Nous avons donc décidé de porter nos premiers efforts sur les registres des États de Bourgogne qui couvrent la période de 1636 à 1787 (cote C 2996 à cote C 3014).

La Bourgogne est un pays d'États : cela signifie que la province, quoique sous l'autorité royale depuis la fin du XV^e siècle, est cogérée par les trois États (clergé, noblesse et tiers-état) jusqu'en 1790. Ces États sont la préfiguration du Conseil régional. Ils se réunissent tous les trois ans à Dijon, sous la présidence du gouverneur de la province, qui est, depuis 1631, le prince de Condé. Entre deux sessions, les Élus des États cogèrent la province avec le pouvoir royal. À partir de 1703, ils siègent dans le Palais des États, devenu en 1831 la Mairie de Dijon. Ils prennent des décrets, consignés sur des registres par des commis aux écritures.

S'ensuivent les Termes. Sa Majesté veut qu'il

«s'ensuivent les Termes. SaMajesté veut qu'il»

*Les États ont décrété qu'il sera imposé
vingt mille livres pendant la Triennalité prochaine,
conformément aux intentions de Sa Majesté pour
continuer à perfectionner les harraes.*

Les États ont décrété qu'il sera imposé
vingt mille livres pendant la Triennalité prochaine,
conformément aux intentions de Sa Majesté, pour
continuer à perfectionner les harraes.

L'intelligence artificielle appliquée à la reconnaissance de l'écriture manuscrite : un long processus d'apprentissage facilité par une équipe de paléographes bénévoles.



Après avoir appris la signification de 20 000 lignes ...



... l'intelligence artificielle est capable de comprendre des textes qu'elle n'a jamais vus.



L'intelligence artificielle (IA) regroupe l'ensemble des techniques visant à simuler l'intelligence humaine par une machine. Au stade actuel de nos connaissances, ces machines sont souvent très spécialisées et nous sommes loin de voir une IA passer le test de Turing, c'est-à-dire atteindre un niveau tel qu'on ne puisse la distinguer d'un humain.

L'une des techniques algorithmiques de l'intelligence artificielle est le réseau de neurones artificiels. Les premiers réseaux de neurones artificiels datent des années 1950. Néanmoins, il a fallu attendre les années 2010 pour que la puissance de calcul des ordinateurs et la capacité à traiter d'immenses bases de données (*Big Data*) soient d'une performance suffisante pour permettre l'apprentissage dit "profond" (*Deep Learning*) indispensable à leur apprentissage.

S'il suffit à un enfant de voir un chat pour en reconnaître instantanément des millions, une intelligence artificielle, un réseau de neurones, aura besoin de milliers, voire de millions de photos de chats pour en reconnaître un.

Ainsi, pour la reconnaissance de l'écriture manuscrite ou *Handwriting Text Recognition* - HTR - , pour chaque ligne on apprend le texte original et sa transcription au réseau de neurones. Pour un apprentissage correct, cela nécessite près de 20 000 lignes que les services des Archives départementales, grâce à l'aide de contributeurs bénévoles formés à la paléographie, ont dû transcrire en fichier texte pour nourrir la machine.

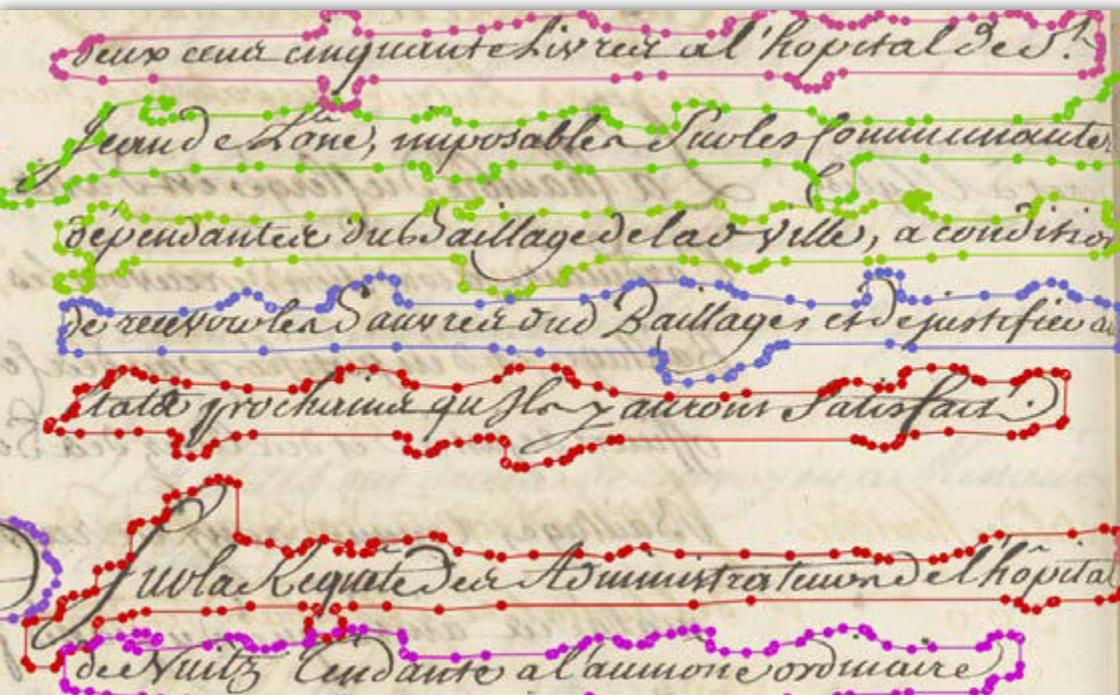
Après la phase d'apprentissage, l'intelligence artificielle est capable de comprendre des textes qu'elle n'a jamais vus.

L'application d'outils à la pointe de la recherche ont conduit à des premiers résultats très encourageants qui vont faire l'objet de publications scientifiques.

Le processus de transcription automatique se découpe en deux parties : la segmentation des lignes puis la transcription de ces lignes.

Pour la phase de segmentation nous avons choisi d'utiliser mask-rcnn. Cet algorithme de réseau de neurones artificiels permet, à l'aide de masques, d'isoler chaque ligne.

La segmentation qui consiste en un détournement de ligne est une étape importante et qui peut engendrer de nombreuses erreurs. L'approche avec mask-rcnn est très efficace et nous a permis une détection de toutes les lignes avec plus de 95% de bonne segmentation.



Support administratif et scientifique assuré par les Archives départementales de la Côte-d'Or

Le suivi administratif et scientifique (numérisation, recherches et expertises paléographiques, transcriptions des documents pour les phases d'apprentissage, relecture et validation des résultats, etc.) sera naturellement assuré par les Archives départementales, sous la responsabilité de leur directeur.

Les Archives départementales s'appuient également sur une équipe de bénévoles, formés à la paléographie, pour la suite des transcriptions en phase d'apprentissage ; la transcription des 20 000 lignes tirées des registres des États de Bourgogne, ont mobilisé un total d'environ 200 jours de travail.

Par ailleurs, il est envisagé de proposer aux étudiants qui seront en master recherche en histoire moderne, dans le cadre du stage de 2 semaines en institution patrimoniale qu'ils doivent accomplir durant leur cursus, de participer au projet « Lettres en Lumières ».

Le budget administratif et scientifique pour les Archives départementales est évalué à 45 000 € sur 3 ans (15 000 € par an).

Équipement informatique

Les apprentissages des différents réseaux de neurones artificiels ont été effectués sur le serveur informatique du LEAD. Ce serveur, qui est déjà équipé de cartes GPU puissantes, est partagé avec les autres chercheurs du LEAD. Afin de donner plus de bande passante dans l'utilisation de ce serveur, il conviendra de le compléter avec des composants dont le montant total est estimé à 40 000 €.

Par ailleurs, l'affichage des images de textes nécessite un terminal (PC) équipé d'un écran de grand format. Le budget estimé pour ce poste de travail est de 5 000 €.

L'IA, c'est quoi ?

Conseil de l'Europe - coe.int – Consulté le 10 novembre 2023

Une définition complexe à opérer et à partager

Si le terme d'« intelligence artificielle » (IA) est entré dans le langage commun et son utilisation devenue banale dans les médias, il n'en existe pas réellement de définition partagée.

Au sens large, le terme désigne en effet indistinctement des systèmes qui sont du domaine de la pure science-fiction (les IA dites « fortes », dotées d'une forme conscience d'elles-mêmes) et des systèmes déjà opérationnels en capacité d'exécuter des tâches très complexes (reconnaissance de visage ou de voix, conduite de véhicule – ces systèmes sont qualifiés d'IA « faibles » ou « modérées »).



Cette confusion contribue à entretenir des craintes purement spéculatives (des IA autonomes conscientes s'opposant aux humains) qui resteraient anecdotiques si elles ne troublaient pas la mesure des réels enjeux, tels que l'impact sur les droits fondamentaux de processus de prise de décision fondés sur des modèles mathématiques, et donc difficile l'élaboration de cadres de régulation.

Une définition à restreindre, au cas par cas, aux technologies utilisées

L'IA est en réalité une discipline jeune d'une soixante d'années, qui réunit des sciences, théories et techniques (notamment logique mathématique, statistiques, probabilités, neurobiologie computationnelle et informatique) et dont le but est de parvenir à faire imiter par une machine les capacités cognitives d'un être humain.

Les spécialistes préfèrent en général employer le nom exact des technologies concrètement en œuvre (qui relèvent aujourd'hui essentiellement de l'apprentissage automatique – machine learning) et sont parfois réticents à employer le terme d'« intelligence » car les résultats, bien qu'extraordinaires dans certains domaines, demeurent encore modestes au regard des ambitions entretenues.

Qu'est-ce que l'apprentissage automatique (*machine learning*) ?

Après avoir connu deux périodes de fort développement (entre 1940 et 1960, puis entre 1980 et 1990), l'IA a connu un nouvel essor en 2010 grâce aux algorithmes dits d'apprentissage automatique. Deux facteurs sont à l'origine de ce nouvel engouement des chercheurs et des industries informatiques : l'accès à des volumes massifs des données et la découverte de la très grande efficacité des processeurs de simples cartes graphiques d'ordinateur pour accélérer le calcul des algorithmes d'apprentissage.

L'actuelle « révolution » de l'IA ne vient donc pas d'une découverte de la recherche fondamentale mais de la possibilité d'exploiter avec efficacité des fondements relativement anciens, tels que l'inférence bayésienne (XVIII^{ème} siècle) ou les neurones formels (1943) pour l'une des sous-classes de l'apprentissage automatique, l'apprentissage profond (ou deep learning).

L'apprentissage automatique a opéré un changement complet de paradigme par rapport à la précédente génération d'IA, les systèmes experts, avec une approche voulue comme inductive : il ne s'agira plus pour un informaticien de coder les règles à la main mais de laisser les ordinateurs les découvrir par corrélation et classification, sur la base d'une quantité massive de données. Autrement dit, l'objectif de l'apprentissage automatique n'est pas réellement d'acquérir des connaissances déjà formalisées mais de comprendre la structure de données et de l'intégrer dans des modèles, notamment pour automatiser des tâches.

Systemes experts et apprentissage automatique : deux conceptions différentes de l'IA



Très concrètement, dans sa phase d'apprentissage, la machine va rechercher les liens entre des données préalablement sélectionnées pour un domaine spécifique (par exemple, sur 10 années, dans différentes villes, le nombre de crèmes glacées vendues et la température de l'air) et les catégoriser. Ce modèle pourra ensuite être utilisé pour

résoudre des questions du type : s'il fait 25°, combien de glaces puis-je espérer vendre dans tel lieu ?

Bien que certains systèmes construisent des modèles de manière relativement autonome, l'intervention humaine demeure encore essentielle, qu'il s'agisse de choisir les données d'apprentissage, d'identifier leurs éventuels biais ou alors, quand cela est possible, de distinguer parmi les corrélations celles pouvant être réellement la cause d'un phénomène (si l'on vend beaucoup plus de glaces pour un lieu donné, est-ce à cause de la température ou de la présence d'un très bon glacier ?).

Le futur de l'IA et ses enjeux

D'après certains experts comme Yann LeCun, chercheur en IA et pionnier de l'apprentissage profond, l'ambition de parvenir à imiter une cognition humaine (ou même animale) nécessiterait de nouvelles découvertes en recherche fondamentale et non une simple évolution des technologies actuelles d'apprentissage automatique. De telles technologies, qui relèvent essentiellement de la mathématique et de la statistique, ne sont en effet pas en mesure d'agir par intuition ou de modéliser rapidement leur environnement.

Les impacts sociétaux, éthiques et sur les droits fondamentaux ne sont donc pas à construire en craignant que l'apprentissage automatique ne fasse émerger une forme conscience artificielle dans les 10 ou 20 années, mais prévenant les biais, les discriminations, les atteintes à la vie privée, à la liberté d'expression ou de conscience voire à la vie elle-même avec les armes autonomes provenant d'une conception de la société la réduisant à un modèle mathématique.

L'intelligence artificielle pour gouverner les archives nationales américaines

Avec des millions de documents numérisés, les archives nationales des États-Unis représentent une mine d'or d'informations. Encore faut-il avoir emporté son tamis et sa pioche. Les utilisateurs reconnaissent en effet avoir du mal à trouver les éléments recherchés. Et le balisage par métadonnées, associé au système de recherche actuellement en vigueur ne fonctionne pas très bien.

Actualitté.com – 20-04-2021 - Clément Solym

La National Archives and Records Administration détient tous les documents gouvernementaux, soit des millions d'œuvres numérisées. Pour répondre aux exigences et impératifs, l'agence a récemment organisé une journée d'information virtuelle. Cette dernière précise comment l'intégration de l'intelligence artificielle et du machine learning permettront de simplifier la vie des usagers.

En effet, des modèles de personnalisation des fonctions de recherche, mais également de catalogage, ou encore l'automatisation du marquage de métadonnées apporteront des solutions plus immédiates. A ce jour, 120 millions de documents archivés, doublés de données de classification, sont présentés dans la base globale. Et l'IA interviendrait pour combler le vide entre les attentes des utilisateurs et les enregistrements.

L'idée est de distinguer, pour exemple, une recherche portant sur Truman, le président et Truman, le porte-avions. Une discrimination que l'outil ne permet actuellement pas d'opérer - on peine à le croire.

Associé à un processus de marquage des métadonnées, pour éviter de se fier aux seules informations inscrites à la main, l'outil deviendrait plus performant. Les algorithmes viendront encadrer au mieux les pratiques, profitant de requêtes effectuées par les internautes, qui se coupleront aux résultats. Le crowdsourcing cumulé à l'intelligence artificielle, autant de belles promesses.

via [Nextgov](#)

L'intelligence artificielle au service du traitement des archives

Utilisée aux Archives nationales depuis 2015, l'intelligence artificielle permet de traiter les fonds de manière inédite, principalement grâce à la reconnaissance d'écritures manuscrites, au bénéfice des archivistes comme des lecteurs.

L'intelligence artificielle, en plus d'être une notion de plus en plus divulguée, est devenue un outil à part entière dans les administrations pour tirer parti des vastes réservoirs de données qu'elles produisent. Le monde de la recherche, de la culture et du patrimoine est lui aussi concerné par ce mouvement de fond : désormais prédomine le sentiment que le stade de l'expérimentation est dépassé pour entrer dans une nouvelle phase des pratiques professionnelles. Aux Archives nationales, c'est à partir de 2015 que l'intelligence artificielle a été employée pour traiter les fonds de manière inédite, principalement en ayant recours à la reconnaissance d'écritures manuscrites, ou *handwritten text recognition* (HTR). Cette technologie s'appuie sur les processus de *machine learning*, qui consistent, à partir de jeux de données, à entraîner une machine à effectuer des actions humaines : ici, en l'occurrence, la transcription de textes manuscrits. Cette technologie répond au rêve longtemps caressé de pouvoir traiter les documents à l'instar de ce qu'était capable d'accomplir l'OCR sur les imprimés. Elle a frayé plusieurs pistes pour le traitement, la diffusion et l'utilisation des archives.

ACCÉDER AU TEXTE DES ARCHIVES : HIMANIS ET ENDP

Lancé fin 2014 par l'IRHT (Institut de recherche et d'histoire des textes) sous l'égide de Dominique Stutzmann, le projet HIMANIS¹ présenta la singularité de choisir comme « terrain de jeu » les registres de la chancellerie royale des XIII^e et XV^e siècles conservés dans le Trésor des chartes. Si l'HTR était alors couramment utilisé sur les écritures contemporaines, il n'avait jamais été employé pour transcrire des écritures si anciennes, comportant de nombreuses abréviations. Le corpus à traiter, soit 199 registres pour plus de 83 000 pages, était en outre suffisamment volumineux pour évaluer la maturité de la technologie. Les premiers résultats ont été obtenus rapidement grâce aux images déjà disponibles au début du projet et aux éditions électroniques d'actes royaux qui fournissaient une « vérité de terrain »², c'est-à-dire une transcription d'une partie du corpus assez exacte pour entraîner l'intelligence artificielle à comprendre

l'écriture et ses mécanismes d'abréviations. Les données obtenues sont des lemmes alignés sur leur image d'origine et sont interrogeables *via* une interface de recherche³.

Le chercheur a la possibilité de retrouver directement au cœur du document les termes de son choix et d'effectuer aussi bien des études d'ordre historique que philologique en s'appuyant sur la statistique lexicale. Du point de vue de l'archiviste, ce moteur de recherche pallie l'absence d'inventaire complet de ces registres. On notera néanmoins qu'un tel outil ne peut remplacer un travail d'indexation classique. Face à la matière brute du texte, c'est à l'utilisateur final de retrouver les termes du Moyen Âge qui traduisent un phénomène ou un concept, de connaître les formes lexicales du latin et de l'ancien français pour retrouver toutes les occurrences pertinentes, d'envisager tout ce qu'une indexation matière pouvait lui suggérer.

Dans la continuité d'HIMANIS, il faut aussi évoquer le projet eNDP, débuté en 2020 et qui porte sur les registres des décisions du chapitre de Notre-Dame de Paris. Autre exemple de coopération entre institutions de conservation et structures de recherche, le but de ce projet, qui a actuellement réussi le stade de la transcription par HTR, vise à explorer le contexte social, économique et urbain dans lequel évoluait le chapitre cathédral⁴.

ACCOMPAGNER LE TRAITEMENT DES ARCHIVES : LECTAUREP ET SIMARA

Une autre possibilité d'exploiter l'intelligence artificielle consiste à récupérer, à partir des documents d'archives, des informations afin de les réutiliser dans le travail de description documentaire. LectAuRep, projet financé par le ministère de la Culture, mené par le Minutier central des notaires parisiens et l'INRIA, a été conduit de 2018 à 2021⁵. Il porte sur les répertoires de notaires, qui sont les clés d'accès fondamentales à leurs minutes, en donnant la date, l'objet de l'acte et le nom des parties. L'HTR a permis de transcrire ces informations. Elles fournissent un ensemble de métadonnées utiles aussi bien pour décrire les documents que pour se prêter à des explorations statistiques. L'équipe est aussi allée plus loin en ayant recours à la technologie de

[1] L'acronyme signifie : « *Historical MANuscripts Indexing for user-controlled Search* ». Carnet de recherche : www.himanis.hypotheses.org

[2] Notamment P. Guérin et L. Celier, *Recueil des documents concernant le Poitou contenus dans les registres de la chancellerie de France*, 14 vol., Poitiers, 1881 ; édition électronique par l'École des chartes : www.corpus.enc.sorbonne.fr/actesroyauxdupoitou

[3] Hébergement par Huma-Num : www.himanis.huma-num.fr/app

[4] Présentation du projet : <https://lamop.hypotheses.org>

[5] LectAuRep signifie : lecture automatique de répertoires. Carnet de recherche du projet : www.lectaurep.hypotheses.org

Archivistique à l'ère de l'IA: Opportunités, défis et besoin d'utilisation responsable

Arbido.ch – 2023/1 - Basma Makhoulf Shabou - Responsable du Master Sciences de l'information à la Haute école de gestion de Genève.

L'archivistique est une discipline qui s'occupe de la gestion, de la préservation et de l'accès aux archives, c'est-à-dire aux documents produits ou reçus par une organisation ou une personne physique dans le cadre de ses activités. À l'ère de l'intelligence artificielle (IA), l'archivistique est confrontée à de nouveaux défis et de nouvelles opportunités.

D'une part, l'IA peut être utilisée pour faciliter la gestion des archives. Elle peut aider à trier, classer et indexer les documents de manière automatique, en utilisant des techniques de traitement automatique du langage naturel (TALN) et d'apprentissage automatique. Elle peut également être utilisée pour détecter et corriger les erreurs dans les descriptions des archives, pour établir des liens entre des documents dispersés dans différents fonds et pour faciliter la recherche et la récupération des documents.

D'autre part, l'IA pose également des défis pour l'archivistique. Tout d'abord, les algorithmes d'IA peuvent être biaisés, car ils sont entraînés sur des données historiques qui reflètent les préjugés et les discriminations de la société. Par conséquent, les archives produites par des systèmes d'IA peuvent perpétuer ces biais. De plus, l'IA peut rendre plus difficile la détermination de l'authenticité et de l'intégrité des documents, car elle peut être utilisée pour créer des documents falsifiés.

Des risques pour la protection de la vie privée

L'IA peut par ailleurs entraîner des conséquences sur la vie privée et la protection des données. L'utilisation de techniques de reconnaissance d'écriture manuscrite, de reconnaissance faciale et de reconnaissance de voix peut révéler des informations personnelles sensibles, telles que l'identité, l'origine ethnique ou les opinions politiques, qui doivent être protégées conformément aux lois sur la protection des données.

En somme, l'archivistique à l'ère de l'IA est confrontée à des opportunités et des défis. Il est important que les archivistes travaillent en étroite collaboration avec les experts en IA pour maximiser les avantages de cette technologie tout en minimisant les risques. Les archivistes doivent également être conscients des implications éthiques et

juridiques de l'utilisation de l'IA dans la gestion des archives et veiller à ce que les droits des individus et de la société soient protégés.

Les avis demeurent partagés à cet égard. Entre défenseurs et résistants, les archivistes tentent de comprendre des différentes retombées de l'adoption des outils utilisant l'IA afin de se positionner rationnellement par rapport à la pertinence et la possibilité d'intégrer ou non les technologies de l'IA pour améliorer et accélérer les différents processus et méthodes de traitement archivistique.

Dans ce cadre, la journée professionnelle de l'AAS était particulièrement riche d'idées et de discussions. Sous forme de symposium, des intervenants des sphères académiques, professionnelles et associatives de Suisse, d'Europe et de l'Amérique du Nord ont été conviés afin de célébrer les cent ans de l'AAS et enrichir ces discussions.

Concepts et pratiques archivistiques dans le cadre de l'IA

Le symposium a commencé par l'intervention de Corinne Rogers, coordinatrice du projet. Après une mise en contexte terminologique et scientifique des concepts importants du projet, elle a souligné les principaux enjeux critiques des caractéristiques et des impératives d'une archive numérique authentique. Elle a également rappelé la nécessité pour les archivistes de façonner activement le développement des technologies d'IA afin de s'assurer qu'elles s'alignent sur les principes et les valeurs archivistiques. Les archivistes sont appelés donc à investir dans une formation continue soutenue et un développement professionnel pour suivre le rythme de l'évolution rapide du paysage technologique. (...)

Préoccupations éthiques

Un autre thème clé du symposium était les considérations éthiques entourant l'utilisation de l'IA dans les archives. Les panélistes ont discuté des risques et des avantages potentiels de l'utilisation des technologies d'IA dans le travail d'archivage, y compris les questions liées à la confidentialité, aux préjugés et à la transparence. Ils ont également souligné la nécessité pour les archivistes d'être proactifs pour répondre à ces préoccupations éthiques et d'élaborer des politiques et des lignes directrices qui garantissent l'utilisation responsable de l'IA dans les archives.

Avec l'avènement de l'intelligence artificielle (IA), l'archivistique est confrontée à de nouveaux enjeux éthiques. L'IA peut être utilisée pour faciliter la gestion des archives, notamment en automatisant certaines tâches répétitives, en identifiant des informations pertinentes dans des documents, ou en aidant à la classification des archives. Cela peut permettre de gagner du temps et de l'efficacité dans le traitement des archives.

Cependant, l'utilisation de l'IA dans l'archivistique soulève également des préoccupations éthiques importantes. Tout d'abord, l'IA peut entraîner des biais dans la classification et la sélection des documents, si elle est mal conçue ou si elle est basée sur des données qui sont elles-mêmes biaisées. Par exemple, l'IA pourrait être utilisée pour sélectionner des documents pour la numérisation ou la conservation, mais si elle est basée sur des critères discriminatoires, elle pourrait reproduire des inégalités et des préjugés. De plus, l'utilisation de l'IA dans l'archivistique pose des questions de confidentialité et de protection des données personnelles. Les données stockées dans les archives peuvent contenir des informations sensibles sur des personnes ou des organisations, et leur utilisation par des algorithmes d'IA peut mettre en danger la vie privée de ces personnes.

Compte tenu de l'opacité de son fonctionnement exact, l'IA peut rendre difficile la vérification et la validation des décisions prises par les systèmes d'archivage automatisés. Les décisions prises par l'IA peuvent être difficiles à comprendre ou à justifier, ce qui peut rendre difficile l'identification et la correction des erreurs. Dans l'ensemble, l'utilisation de l'IA dans l'archivistique doit être guidée par des principes éthiques solides, en veillant à ce que les avantages de l'automatisation soient équilibrés avec la protection des droits des individus et des organisations.

IA : « On va avoir de plus en plus de mal à constituer des archives authentiques »

Emilie Echaroux - usbeketrica.com - 26 septembre 2023

Notre patrimoine se dématérialise à grands pas. En 2020, près de 27 260 Gigaoctets d'archives numériques ont été collectés par le ministère de la Culture, rejoignant la librairie immatérielle des Archives Nationales. Mais à quoi ressemblera notre mémoire numérique collective dans le futur, si de nouveaux contenus sont générés automatiquement avec le recours à l'intelligence artificielle ? Le philosophe et informaticien Jean-Gabriel Ganascia nous répond. (...)

L'un des risques identifiés est la multiplication de contenus fallacieux. Selon un rapport d'Europol, 90 % du contenu en ligne pourrait être généré artificiellement d'ici à 2026, et ce sans forcément de supervision humaine. Doit-on s'attendre à ce que les archives du futur soient fondées sur des fake news ?

JEAN-GABRIEL GANASCIA :

Nos mémoires numériques sont aussi vertigineuses que fragiles. Et pour cause : leurs supports sont particulièrement faillibles. Dès qu'une petite impureté les entache, aussi microscopique soit-elle, les données stockées sur ces supports deviennent illisibles.

Notre mémoire numérique est également fragilisée par la présence d'informations fausses.

Si les fake news ont toujours existé, on en trouve aujourd'hui dans une proportion démesurée, alourdissant la tâche de vérification des historiens et des journalistes. La génération de contenus par le biais de l'intelligence artificielle décuple ce problème.

À l'avenir, les fausses nouvelles vont non seulement subsister, mais aussi se multiplier. Et il se peut que d'ici dix ans, il y ait plus de contenu généré par le biais de l'intelligence artificielle que par des humains, et qu'il y ait proportionnellement plus de fake news générées que de vraies informations.

De fait, on va avoir de plus en plus de mal à constituer des archives authentiques de notre futur, ce qui risque de mettre à mal notre mémoire collective.

Devons-nous d'ores et déjà imaginer de nouvelles manières de conserver notre mémoire ?

JEAN-GABRIEL GANASCIA :

Il s'agit de développer des techniques permettant de détecter les informations générées par l'intelligence artificielle. Il existe plusieurs solutions en la matière, dont le watermarking, ou filigrane en français.

De la même manière que sur un billet de banque, une marque imperceptible à l'œil nu [semblable à un tatouage numérique, nldr] permet l'identification et l'authentification d'un document, le préservant de toute copie malintentionnée.

Cette technique fonctionne très bien lorsqu'il y a une forte redondance d'information, comme avec les images ou la musique, mais ce n'est pas aussi évident avec du texte.
(...)



[Emilie Echaroux](#)

- 26 septembre 2023

Recommandation sur l'éthique de l'intelligence artificielle (extraits) - UNESCO - Novembre 2021

DOMAINE STRATÉGIQUE 7 : CULTURE

- 94.** Les États membres sont encouragés à établir, s'il y a lieu, des systèmes d'IA dans les domaines de la conservation, de l'enrichissement, de la compréhension, de la promotion, de la gestion et de l'accessibilité du patrimoine culturel matériel, documentaire et immatériel, en particulier les langues en péril et les langues et savoirs autochtones, en mettant en place ou en actualisant par exemple les programmes éducatifs concernant l'application de systèmes d'IA dans ces domaines, selon les besoins et suivant une approche participative, à l'intention des institutions et du public.
- 95.** Les États membres sont encouragés à examiner et à traiter les répercussions des systèmes d'IA sur la culture, en particulier celles des applications de traitement du langage naturel, comme la traduction automatique et les assistants vocaux, sur les nuances de l'expression et du langage humains. Ces évaluations devraient permettre de contribuer à la conception et à la mise en œuvre de stratégies visant à tirer parti au maximum des avantages de ces systèmes en réduisant les inégalités culturelles et en améliorant la compréhension humaine, ainsi qu'en luttant contre leurs effets négatifs, tels qu'un recul de l'usage qui pourrait entraîner la disparition de langues en péril, de dialectes locaux et de variations tonales et culturelles associées à l'expression et au langage humains. (...)

- 100.** Les États membres devraient encourager les musées, les galeries, les bibliothèques et les archives à l'échelle nationale à utiliser des systèmes d'IA afin de valoriser leurs collections et d'enrichir leurs bibliothèques, leurs bases de données et leurs bases de connaissance, tout en facilitant l'accès à leurs usagers.

DOMAINE STRATÉGIQUE 9 : COMMUNICATION ET INFORMATION

- 112.** Les États membres devraient utiliser les systèmes d'IA pour améliorer l'accès à l'information et au savoir. Ils peuvent notamment mettre en place des initiatives visant à aider les chercheurs, le milieu universitaire, les journalistes, le grand public et les développeurs, à améliorer la liberté d'expression, les libertés académiques et scientifiques ainsi que l'accès à l'information, et à accroître la divulgation proactive des données et informations officielles. (...)
- 114.** Les États membres devraient investir dans les compétences numériques et l'éducation aux médias et à l'information et les promouvoir, afin de renforcer la pensée critique et les compétences indispensables pour comprendre l'utilisation et les implications des systèmes d'IA, en vue d'atténuer et de combattre la désinformation, la mésinformation et le discours de haine. Une meilleure compréhension et une meilleure évaluation des effets positifs et éventuellement indésirables des systèmes de recommandation devraient faire partie de ces efforts.
- 115.** Les États membres devraient créer des environnements favorables aux médias, dans lesquels ces derniers disposeraient des droits et ressources nécessaires pour rendre compte efficacement des avantages et des inconvénients des systèmes d'IA, et encourager les médias à faire un usage éthique de ces systèmes dans leur travail.

L'INTELLIGENCE ARTIFICIELLE EST-ELLE UNE MENACE OU UNE CHANCE POUR LES PROFESSIONNELS EN CHARGE DE L'ARCHIVAGE DES DONNÉES DES ENTREPRISES ET ORGANISATIONS ?

blog.cr2pa – CR2PA club de l'archivage managerial - 15 juillet 2019

Les outils d'IA disponibles pour le processus d'archivage vont-ils faciliter ce dernier ou le faire disparaître ? Comment l'IA peut-elle permettre de mieux archiver ? Qu'ont en commun archivage managerial et IA ? L'IA redéfinit-elle ce que l'on va archiver demain ?

A l'heure où le tsunami numérique impacte notre façon de penser l'archivage, de nouveaux outils sont ou seront bientôt disponibles pour nous aider à archiver...

Face, d'un côté, à l'infobésité et, de l'autre, aux progrès de l'intelligence artificielle, comment se positionnent les managers de la donnée que nous sommes devenus ?

Sur ce point, les convictions du CR2PA ne perdent rien de leur pertinence. Au contraire, on peut même avancer qu'elles se trouvent renforcées.

Au premier rang desquelles, celle que l'humain reste au centre du management de l'information et que, en quelque sorte, l'intelligence de l'IA est celle que nous voulons bien lui donner. Dans ces conditions, comment l'IA peut-elle aider sur tout le cycle de vie de la donnée et quelles valorisations peut-elle apporter (information, connaissance dans un contexte) en contribuant aux décisions directes ou indirectes ?

Fondamentalement, les technologies d'IA et d'apprentissage automatique reproduisent la cognition et les capacités humaines d'apprentissage, de façon accélérée.

Elles ne sont, par conséquent, opérantes qu'à partir d'une stratégie d'apprentissage saine et de données de qualité. Une première approche est par conséquent de contrôler la qualité des données fournies aux algorithmes d'apprentissage ou d'intégrer aux IA des paramètres régulateurs.

Les managers de la donnée devront demain rendre l'IA intelligible à tous ceux qui y sont confrontés et être au cœur de ce processus d'intelligibilité.

Outil d'aide à l'archivage, l'IA peut également -et peut être surtout- 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.

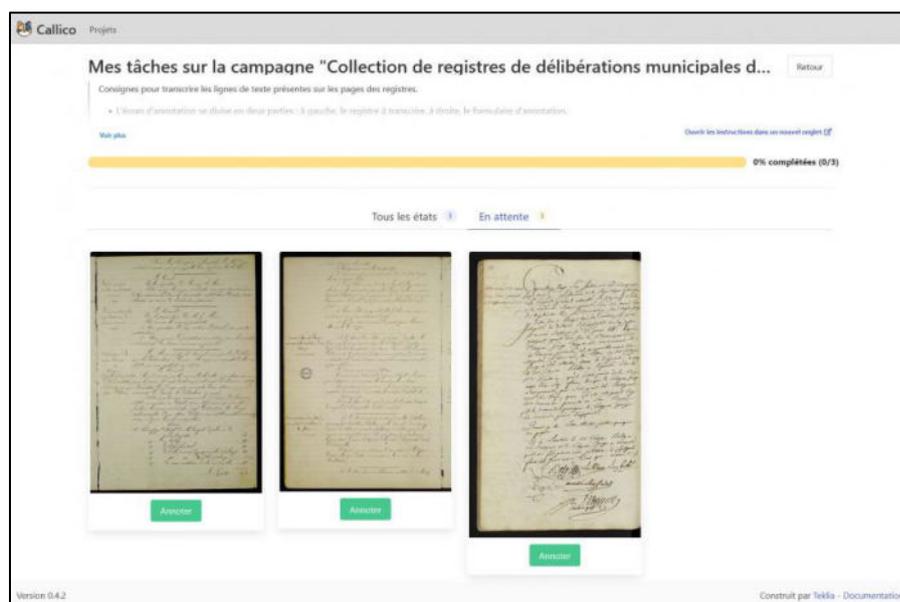
Elle pourra permettre une meilleure traçabilité et contribuer à une gestion plus efficace de l'archivage : reconnaître que la donnée est engageante et doit être archivée, apporter de la « distance » par rapport aux éléments du référentiel d'archivage, aider à la classification dans le référentiel, en aval, gérer les comptes rendus d'intervention, accéder de façon directe au bon élément du référentiel de conservation, ou même permettre une meilleure acculturation des collaborateurs à l'archivage...

Alors, archivage et Intelligence artificielle même combat ?

Oui si l'on part du postulat que l'humain prime sur les outils et qu'il faut rendre l'IA intelligible à tous ceux qui y sont confrontés, oui si tous les acteurs restent conscients des risques, et maîtres du processus de qualification des documents engageants (définir ce qui doit être conservé, pour quelle durée, etc...), qui seul permet une maîtrise dans la durée de l'information à risque.

Éduquez l'intelligence artificielle avec les Archives de Belfort

Rfgenealogie.com - Par Pierre-Valéry Archassal - Territoire de Belfort (90) - 06.03.2023



*Apprendre à lire à la machine pour qu'ensuite elle le fasse à notre place, c'est la première phase du projet lancé par les Archives de Belfort en partenariat avec l'entreprise Teklia. Chaque bénévole se voit attribuer un lot de pages qu'il peut transcrire à son rythme dans la plateforme dédiée.
Crédits Capture AM Belfort*

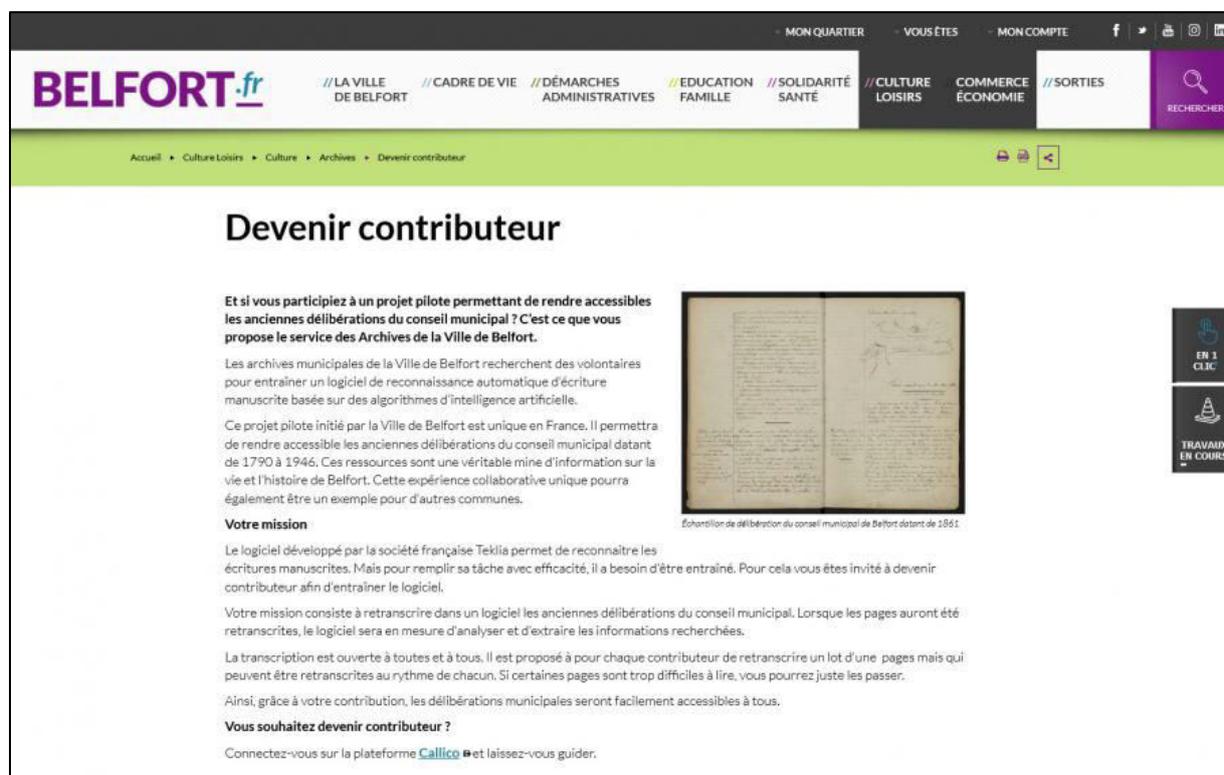
Comme nous l'avons détaillé dans notre dossier intitulé « Intelligence artificielle : quand les machines liront pour nous », les ordinateurs ont besoin d'apprendre l'écriture d'un scribe avant de devenir autonome pour lire tout ce qu'il a rédigé. Cet apprentissage passe, dans la plupart des cas, par un accompagnement humain : plusieurs pages sont transcrites par de « vrais » lecteurs pour servir ensuite de base de comparaison à la machine et lui permettre de décrypter les pages restantes.

Les différentes expérimentations menées dans ce domaine ont aussi montré qu'il était plus simple d'apprendre à une machine à lire un corpus de documents cohérent.

Autrement dit, il est trop tôt pour qu'une intelligence artificielle passe instantanément d'un acte de baptême de XV^e siècle à un testament du XVIII^e siècle ou un courrier de Poilu. Les Archives nationales ont choisi par exemple de cibler leurs efforts en la matière sur les répertoires de notaires, les Archives départementales de la Côte d'Or ont privilégié les registres des décrets des États de Bourgogne et le projet Socface se consacre exclusivement aux listes nominatives de population de 1836 à 1936.

Le projet que lancent aujourd'hui les Archives de Belfort concerne les anciennes délibérations du conseil municipal, qui constituent une source d'informations essentielles pour connaître l'histoire de la ville sur la vaste période qui va de 1790 à 1946. Depuis l'an

dernier, ces précieux registres sont numérisés et, bientôt, ils vont être passés à la moulinette de l'intelligence artificielle pour être intégralement transcrits afin de les rendre plus accessibles.



The screenshot shows the website **BELFORT.fr** with a navigation menu including 'LA VILLE DE BELFORT', 'CADRE DE VIE', 'DÉMARCHES ADMINISTRATIVES', 'EDUCATION FAMILLE', 'SOLIDARITÉ SANTÉ', 'CULTURE LOISIRS', 'COMMERCE ÉCONOMIE', and 'SORTIES'. The main heading is 'Devenir contributeur'. The text invites users to participate in a pilot project to make old municipal council deliberations accessible by training an AI-based handwriting recognition software. It mentions that the project is unique in France and dates back to 1790-1946. A section titled 'Votre mission' explains that users will transcribe pages in a software, and the software will then analyze and extract information. A small image of an old manuscript is shown with the caption 'Échantillon de délibération du conseil municipal de Belfort datant de 1861'. On the right side, there are two buttons: 'EN 1 CLIC' and 'TRAVAUX EN COURS'.

S'inscrire pour participer à distance au projet belfortin concerne tous les internautes car le bénévole fourni servira ensuite à d'autres projets identiques, sur d'autres archives, dans d'autres régions. Crédits Capture AM Belfort

C'est la société française Teklia, celle-là même qui œuvre pour Socface, qui gère le projet en partenariat avec la ville de Belfort. Les logiciels de reconnaissance automatique des écritures manuscrites qu'elle développe s'appuient sur cette fameuse phase d'apprentissage pour laquelle l'intervention humaine est indispensable. Un appel à bénévoles est donc lancé, ouvert à toutes et à tous. Il est proposé à chaque contributeur de retranscrire quelques pages à son propre rythme. Si certaines pages sont trop difficiles à lire, il suffit de les passer, elles seront ensuite transmises à d'autres lecteurs plus aguerris.

Nous encourageons donc chacun d'entre vous, ayant ou non des ancêtres à Belfort, à participer à ce projet qui sera utile à tous car l'apprentissage acquis par la machine sera ensuite transférable à d'autres projets qui, peut-être, vous concerneront directement. S'inscrire est facile, il suffit de créer son compte sur la plateforme Callico (une adresse électronique suffit, pas besoin d'indiquer son nom ni plus de coordonnées) et de suivre les indications données par Teklia. Si le nombre de volontaires est important, l'opération pourrait prendre un à deux mois seulement !