

Faculté de philosophie, arts et lettres

Archives et archivistes à l'ère de l'intelligence artificielle.
Bouleversement ou reconfiguration des pratiques et du
cycle de vie des archives ?

**Déclaration sur l'honneur, mémoires/TFE de la Faculté de philosophie, arts et lettres,
UCLouvain**

Je déclare qu'il s'agit d'un travail original et personnel, que toutes les sources référencées ont été indiquées dans leur totalité et ce, quelle que soit leur provenance, et que l'utilisation des outils d'intelligence artificielle est conforme aux consignes générales d'utilisation publiées par l'UCLouvain et disponible à cette adresse :

<https://cdn.uclouvain.be/groups/cms-editors-p1/portail/reglement/Consignes-ChatGPT-version-etudiante.pdf?itok=SebXXm87>

Je suis conscient·e que le fait de ne pas citer une source, de ne pas la citer clairement et complètement, de ne pas annoncer dans une partie spécifique en début du mémoire l'utilisation que j'ai faite des outils d'intelligence artificielle ou de les utiliser de manière contraire à ce qui est prévu dans la note institutionnelle constituent des irrégularités graves au regard du Règlement des études et des examens de l'Université catholique de Louvain (Chapitre 4, section 7, article 107 à 114) au sein de l'Université. J'ai notamment pris connaissance des risques de sanctions académiques et disciplinaires encourues.

Au vu de ce qui précède, je déclare sur l'honneur ne pas avoir commis de plagiat ou toute autre forme de fraude et de n'avoir pas utilisé les outils d'intelligences artificielles en dehors de ce qui est accepté à l'UCLouvain.

Nom, Prénom : BERNY Camille

Date : 26/05/2026

Signature de l'étudiant·e :



Introduction

Dans les années 1960, Yves Pérotin propose, dans un article publié en 1961 dans *Seine et Paris*¹, une présentation des archives en trois étapes successives (archives courantes, intermédiaires et définitives) qui « visent la gestion spatio-temporelle des archives à travers la fixation à l'avance des délais de conservation et la gestion des espaces d'archivage »². Inspirée de réflexions nord-américaines des décennies précédentes et des travaux de l'archiviste Schellenberg sur les valeurs administratives et historiques³, cette théorie des « trois âges » deviendra l'un des cadres les plus largement mobilisés dans l'espace francophone pour organiser le traitement des archives⁴.

La version proposée par Pérotin ne constitue toutefois qu'une interprétation parmi d'autres du cycle de vie. Il convient en effet de rappeler qu'il n'existe pas un cycle de vie unique, mais une pluralité d'interprétations et de déclinaisons selon les contextes nationaux, les cultures institutionnelles ou les cadres réglementaires⁵. Dans d'autres traditions archivistiques, notamment nord-américaines, le modèle se fonde sur une séparation plus nette entre la gestion courante et le traitement archivistique, tandis que les pratiques belges, françaises, québécoises ou encore espagnoles adoptent des découpages divers selon les besoins organisationnels et les cadres juridiques⁶. Dans les faits, plusieurs auteurs montrent que ces frontières sont souvent moins franches que ne le laisse entendre la théorie, et que les documents passent d'un état à l'autre de manière plus progressive qu'idéalement décrite⁷. à ce propos, il est donc nécessaire

¹ PÉROTIN Yves, « L'administration et les «trois âges» des archives », dans *Seine-et-Paris*, n°20, octobre 1961, p.31-33.

² Ainsi, « les archives courantes, du premier âge encore fréquemment utilisées, sont conservées à proximité de leurs producteurs dans les bureaux. Les archives intermédiaires, du deuxième âge qui ne sont pas d'usage courant, sont conservées dans des locaux de préarchivage et enfin les archives définitives/historiques dont l'utilité administrative est achevée, mais qui présentent un intérêt de recherche et patrimonial sont versées au service des archives ».

SAAD Nora, ELMQADDEM Nouredine, VASSEUR Edouard, « Archivage à l'ère du numérique : objet et cycle de vie », dans *Journal of Information Sciences*, vol.21, n°2, 2022, p.180.

³ « La valeur primaire ou première est associée à la fonction du document pour celui qui le produit, au sein d'une activité et de preuve administrative, légale ou financière, c'est-à-dire les raisons pour lesquelles le document a été produit. La valeur secondaire ou seconde est associée à la fonction de témoignage et d'information ».

GUYON Céline, « Le temps long de l'archive », *Séminaire de l'Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB)*, juin 2019, [tps://hal.science/hal-04371293v1](https://hal.science/hal-04371293v1), consulté le 19 juin 2025.

⁴ KERN Gilliane, HOLGADO Sandra, COTTIN Michel, « Cinquante nuances de cycle de vie. Quelles évolutions possibles ? », dans *Les cahiers du numérique*, n°2, 2015, p.39.

⁵ Cf. *Ibid.*, p.37-76.

⁶ *Ibid.*, p.41.

⁷ Par exemple ROUSSEAU Jean-Yves et COUTURE Carol (éd.), *Les fondements de la discipline archivistique*, Québec, Presses universitaires du Québec, 2008, SAAD Nora, ELMQADDEM Nouredine, VASSEUR Edouard, « Archivage à l'ère du numérique : objet et cycle de vie », dans *op.cit.*, p.173-193.

de rappeler que le terme « archivistique » peut recouvrir différentes réalités, ou renfermer différentes acceptions : d'une part, par une approche administrative, centrée sur la valeur primaire des documents, d'autre part par une approche plus traditionnelle, tournée vers leur valeur secondaire, ou encore par une « approche intégrée », qui cherche à tenir ensemble ces deux dimensions et à tenir compte de la continuité des usages tout au long de la vie documentaire⁸. C'est dans cette dernière approche que s'inscrit ce travail, car elle permet d'articuler de manière cohérente les transformations successives introduites par le numérique, puis par l'intelligence artificielle.

À partir des années 1990, dans le cadre de la pensée postmoderne, qui souligne que les archives ne sont jamais des ensembles neutres, mais des constructions sociales façonnées par les interventions successives des producteurs, des archivistes et des usagers, plusieurs travaux commencent à interroger la linéarité implicite du cycle de vie (par exemple ceux de Brothman, Cook, ou Ketelaar)⁹. Parallèlement à cela, l'essor du numérique modifie progressivement les conditions d'existence et de gestion des documents, qu'il s'agisse de leur forme, de leur circulation ou des dispositifs nécessaires à leur traitement. L'ensemble de ces évolutions conduit à questionner les modèles historiquement fondés sur une succession d'étapes¹⁰ et à envisager des lectures plus continues des processus documentaires, sans pour autant rendre caducs les cadres existants. Parmi ces modèles alternatifs, le *records continuum*, développé dans les années 1990 par l'archiviste australien Frank Upward et ses collègues, cherche à rendre compte des contextes complexes dans lesquels les documents sont créés et gérés à l'ère du numérique. Contrairement au modèle classique du cycle de vie, il ne considère plus les archives comme des documents passant successivement par plusieurs âges distincts, mais comme des documents dont les dimensions de création, de gestion, de conservation et d'usage coexistent de manière continue et simultanée. Cette approche invite ainsi à repenser les temporalités archivistiques dans les environnements numériques.

De son côté, l'IA n'est pas neuve. Dès 1950, Alan Turing s'interroge sur la possibilité d'un comportement « intelligent » de la machine, puis, en 1955, le *Dartmouth Summer Research Project* marque la naissance officielle du domaine. Les années 1990, et la victoire du

⁸ ROUSSEAU Jean-Yves et COUTURE Carol, « La place de l'archivistique dans la gestion de l'information », dans ROUSSEAU Jean-Yves et COUTURE Carol (éd.), *op.cit.*, p.50.

⁹ Voir LEMAY Yvon et KLEIN Anne, « Les archives définitives : un début de parcours. Revisiter le cycle de vie et le Records continuum », dans *Archivaria*, n°77, 2014, p.73-102.

¹⁰ L'une des critiques les plus connues est celle de Marcel Caya, en 2004. CAYA Marcel, « La théorie des trois âges en archivistique. En avons-nous toujours besoin ? », conférence donnée à l'Ecole nationale des chartes, le 2 décembre 2004, <http://elec.enc.sorbonne.fr/conferences/caya> (consulté le 15 janvier 2025).

superordinateur Deep Blue contre le champion d'échec Kasparov, relancent l'intérêt pour l'IA et préparent l'essor de l'apprentissage profond à partir des années 2010. Depuis lors, les progrès rapides dans le traitement du langage, la vision par ordinateur et les systèmes autonomes, ainsi que l'apparition des grands modèles de langage en 2023, puis des modèles multimodaux en 2024, ont rendu ces technologies plus accessibles et plus présentes dans de nombreux secteurs¹¹. « C'est à partir de 2015 que l'intelligence artificielle a commencé à être véritablement utilisée pour traiter des fonds d'archives. Notamment, deux technologies, la reconnaissance des écritures manuscrites (ou HTR) et la reconnaissance d'entités nommées (NER), contribuent à renouveler profondément le traitement des archives, mais aussi leurs modalités de consultation et d'usage »¹². L'apparition de *ChatGPT* en 2022 constitue cependant un tournant dans cette évolution, en contribuant à démocratiser l'accès aux IA génératives accessibles à un très large public et à accélérer leur diffusion dans de nombreux usages professionnels. Après cette phase de généralisation, les développements récents tendent davantage vers une spécialisation des modèles et des outils, adaptés à des tâches et à des domaines spécifiques, notamment dans les environnements documentaires et archivistiques.

À partir de ces évolutions, qui ont déjà profondément marqué la pratique archivistique, ce travail interroge la manière dont l'intelligence artificielle s'inscrit dans cette continuité. Constitue-t-elle une rupture, remettant en cause la logique séquentielle et les catégories classiques de la discipline, ou bien participe-t-elle d'une reconfiguration qui prolonge et accentue les transformations amorcées avec le numérique ? En quoi le numérique, puis l'IA, ont-ils modifié les pratiques et la lecture du cycle de vie des archives ? Comment l'IA intervient-elle dans les opérations, qu'il s'agisse d'automatiser, d'interpréter ou de dériver des informations ? Quelles implications ces traitements ont-ils pour l'agentivité et la responsabilité des archivistes ? Comment l'IA renégocie-t-elle le rapport archivistes-archives ? Les applications de l'IA et les études les concernant sont la plupart du temps « tournées utilisateurs », sans « vision métier »¹³, c'est pourquoi notre propos se concentrera sur la pratique archivistique, du côté des professionnels.

¹¹ « Qu'est-ce que l'intelligence artificielle (IA) », sur *ISO*, <https://www.iso.org/fr/contents/news/insights/AI/what-is-ai-all-you-need-to-know.html>, consulté le 16 juin 2025.

¹² MOUFFLET Jean-François, « L'intelligence artificielle au service de la description et de la diffusion des archives. Retour sur des projets conduits dans le réseau des services d'archives publics », sur *Archivistes*, <https://www.archivistes.org/Journee-d-etude-AAF-Ecole-nationale-des-Chartes-L-intelligence-artificielle> (consulté le 10 janvier 2025).

¹³ Discours d'introduction d'Emmanuel Bodart, directeur des Archives de l'Etat à Namur, lors de *l'International Archives Symposium*, le 5 juin 2025.

Pour y répondre, ce travail s'appuie, outre la littérature scientifique en archivistique¹⁴, sur un corpus diversifié combinant littérature professionnelle, documentation institutionnelle et retours d'expérience issus du terrain. Une attention particulière a été portée à la littérature professionnelle, notamment à travers des revues spécialisées comme *Archimag*, ainsi qu'aux ressources diffusées par des associations professionnelles et institutions archivistiques.

Ce mémoire mobilise également l'analyse de projets de recherche et d'expérimentation liés à l'IA dans le domaine patrimonial et archivistique, à partir de la documentation et des présentations disponibles sur leurs sites web (par exemple, *Himanis* ou *Socface*). Il s'appuie en outre sur des entretiens et correspondances avec des professionnels engagés dans des projets d'IA¹⁵, ainsi que sur une participation à des colloques¹⁶, journées d'étude et webinaires consacrés à l'IA en archivistique, qui ont permis de recueillir des retours d'expérience et des réflexions professionnelles contemporaines. Compte tenu de l'évolution particulièrement rapide des technologies et des publications liées à l'intelligence artificielle, la veille bibliographique et l'état de l'art mobilisés dans ce mémoire ont été arrêtés en mars 2026. L'IA a elle aussi été impliquée dans ce travail dans le cadre d'une démarche réflexive consistant à lui soumettre certaines hypothèses, les challenger ou explorer certaines idées¹⁷.

Les prolégomènes rassemblent les principes nécessaires pour comprendre le fonctionnement des traitements d'IA mobilisés dans les environnements documentaires afin de fournir la grammaire minimale indispensable à leur supervision et à leur analyse. Dans sa première partie, le mémoire examine ensuite les transformations introduites par le numérique, qu'il s'agisse de la nature des documents, des dispositifs de gestion et de conservation, ou encore des gestes professionnels et des modèles théoriques mobilisés pour décrire les opérations. Cette première partie a pour objectif de mettre en évidence la manière dont le numérique a modifié les fondements de la pratique archivistique (au niveau de la forme documentaire, les dispositifs de gestion et de conservation et la pratique professionnelle), et a contribué à questionner la linéarité du cycle de vie¹⁸. Dans un second temps, le mémoire

¹⁴ La littérature scientifique en archivistique, à retrouver en bibliographie, questionne l'influence du numérique sur le cycle de vie des archives, l'influence de l'IA sur la pratique archivistique, mais pas l'influence de l'IA sur le cycle de vie des archives.

¹⁵ Echanges avec Johan Pierret, délégué à l'archivage au Ministère de la Santé et de la Sécurité sociale du Grand-Duché du Luxembourg, 25 juillet 2025, et Amandine Gorse, archiviste à la Chambre des Députés du Grand-Duché du Luxembourg, le 6 juin 2025, à l'*International Archives Symposium*.

¹⁶ *International Archives Symposium* de Namur (5 et 6 juin 2025), Webinaire de l'AAF, « Mieux comprendre l'intelligence artificielle et son utilisation dans les archives » (30 septembre 2025).

¹⁷ Les réflexions menées de concert avec l'IA sont indiquées en notes de bas de page.

¹⁸ Notre idée de départ avait été de questionner l'influence du numérique sur le cycle de vie des archives en suivant la chronologie des trois âges. Ce découpage s'est vite avéré très scolaire, et finalement limité, nous le verrons, du fait de la remise en question du cycle traditionnel et à la mise au point du *records continuum*. En travaillant par

s'attache à comprendre ce que l'intelligence artificielle ajoute à ces évolutions, en analysant ses effets sur la forme documentaire, sur les procédés de traitement et de classification, sur les responsabilités professionnelles ainsi que sur la pertinence respective des modèles du cycle de vie et du *continuum*. Il s'agit de déterminer dans quelle mesure l'IA prolonge les dynamiques amorcées avec le numérique ou si elle introduit des déplacements d'un autre ordre.

âge, il nous est apparu que les changements annoncés formaient une séquence forme-dispositifs-pratiques professionnelle-cycle de vie, qui répondait également aux évolutions constatées suite à l'arrivée de l'IA dans la pratique archivistique.

Prolégomènes. Intelligence artificielle : définitions et fonctionnement

À l'ère du numérique avancé, l'intelligence artificielle s'impose comme un levier de transformation majeur dans de nombreux domaines, y compris celui de l'archivistique. Elle interroge en profondeur le concept traditionnel du cycle de vie des archives, et ouvre la voie à une reconfiguration des théories et pratiques professionnelles. Afin de mieux appréhender les modifications et les enjeux soulevés par l'IA, il nous semble d'abord important de faire un tour d'horizon de cet univers, en définissant les concepts et notions liés à l'IA¹⁹.

1. Définitions de base

Intelligence artificielle (IA)

Terme créé en 1955 par John McCarthy lors de l'école d'été de Dartmouth²⁰, l'IA est désormais définie par le Parlement européen comme étant la « possibilité pour une machine de reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité »²¹. L'IA repose sur l'interdisciplinarité théorique et pratique, mêlant des approches statistiques, numériques, algorithmiques, logiques, ou encore mathématiques.

Selon le Parlement européen, il existe deux types d'IA : les logiciels, qui nous intéresseront plus particulièrement dans ce travail (tels que les « assistants virtuels, logiciels d'analyse d'images, moteurs de recherche, systèmes de reconnaissance faciale et vocale ») et l'IA incarnée (dans des « robots, voitures autonomes, drones, ou l'internet des objets »²²).

Nous parlons généralement d'IA faibles, car elles sont conçues pour accomplir une tâche spécifique, sans conscience, compréhension, ni capacité à généraliser au-delà de cette tâche. Elles ont la capacité d'apprendre, mais en restant limitées à leur domaine spécifique. Il s'agit d'outils de reconnaissance faciale, d'assistants vocaux, d'algorithmes de recommandations ou de traduction automatique. À l'opposé, l'IA forte, qui n'existe pas aujourd'hui, serait capable de raisonner, de comprendre, d'apprendre et de s'adapter comme un humain, dans n'importe quel domaine²³.

¹⁹ Notez que les définitions ne sont pas classées par ordre alphabétique, mais suivent un ordre logique pour la compréhension des concepts.

²⁰ MACLURE Jocelyn et SAINT-PIERRE Marie-Noëlle, « Le nouvel âge de l'intelligence artificielle : une synthèse des enjeux éthiques », dans *Les cahiers de propriété intellectuelle*, vol. 30, n°3, 2018, p. 745 ; HENNO Jacques, « 1956 : et l'intelligence artificielle devint une science », sur *Les Echos*, <https://www.lesechos.fr/2017/08/1956-et-lintelligence-artificielle-devint-une-science-181042>, consulté le 11 juin 2025.

²¹ « Intelligence artificielle : définition et utilisation », sur Parlement européen, <https://www.europarl.europa.eu/topics/fr/article/20200827STO85804/intelligence-artificielle-definition-et-utilisation>, consulté le 28 mai 2025.

²² *Ibid.*

²³ *Ibid.* ; « Intelligence artificielle », dans DataFranca, *Les 101 mots de l'intelligence artificielle. Petit guide du vocabulaire essentiel de la science des données et de l'intelligence artificielle*,

Algorithme

Il s'agit d'une suite d'instructions logiques et mathématiques permettant de résoudre un problème ou d'exécuter une tâche. On pourrait comparer un algorithme à une recette de cuisine. Les algorithmes d'IA, exprimés sous la forme d'un logiciel ou d'une application, permettent par exemple à une machine d'apprendre à partir des données (algorithmes d'apprentissage automatique, tels que les réseaux de neurones, les arbres de décision), de prendre des décisions (algorithmes prédictifs, qui ont pour but de prédire un résultat futur à partir de données passées ou présentes), ou encore de s'adapter et de s'améliorer (algorithmes d'apprentissage par renforcement)²⁴.

Donnée (*data*)

Il s'agit de la plus petite unité d'information brute, porteuse de sens²⁵. C'est une « représentation formalisée de l'information adaptée à la communication, l'interprétation ou le traitement »²⁶. Dans le domaine de l'informatique, il s'agit donc d'une petite unité d'information digitale, que ce soit du texte, ou d'un nombre²⁷.

Les jeux de données sont l'ensemble des données utilisées pour entraîner un modèle.

Système d'intelligence artificielle

« Un système d'IA est un système informatique spécifiquement conçu pour analyser des données, identifier des modèles et utiliser cette connaissance pour générer des décisions ou des prédictions informées »²⁸. Autrement dit, il s'agit d'une application complète qui utilise un ou plusieurs modèles d'IA pour générer des résultats (prédictions, recommandations, décisions, contenu), sur base de règles émises par l'humain et de jeux de données²⁹.

https://datafranca.org/wiki/Intelligence_artificielle, consulté le 28 mai 2025 ; « Intelligence artificielle faible », *ibid.*, https://datafranca.org/wiki/Intelligence_artificielle_faible, consulté le 28 mai 2025.

²⁴ « Algorithme », *ibid.*, <https://datafranca.org/wiki/Algorithme>, consulté le 28 mai 2025 ; « Algorithme », sur *Commission Nationale de l'Informatique et des Libertés (CNIL)*, <https://www.cnil.fr/fr/definition/algorithme>, consulté le 28 mai 2025.

²⁵ « Data », sur *InterPARES Trust AI*, <https://interparestrustai.org/terminology/term/data>, consulté le 12 juin 2025.

²⁶ « Données », dans Services des ressources informationnelles et archives de l'Université de Lausanne, *Politique de records management. Glossaire*, 2014, https://www.unil.ch/files/live/sites/unil/files/02-universite/0204-organisation/020404-unites-et-services/UNIRIS/references/UNIL_POL_RM_Glossaire_VF.pdf, consulté le 13 juin 2025.

²⁷ ROUSE Margaret, « Data », sur *Techopedia*, <https://www.techopedia.com/definition/807/data>, consulté le 13 juin 2025.

²⁸ « Qu'est-ce qu'un système d'intelligence artificielle ? », sur *Autorité de protection des données (APD)*, <https://www.autoriteprotectiondonnees.be/citoyen/themes/intelligence-artificielle/-qu-est-ce-qu-un-systeme-d-intelligence-artificielle->, consulté le 16 juin 2025.

²⁹ « Différence entre un système d'IA et un modèle d'IA », sur *Dastra*, <https://www.dastra.eu/fr/article/difference-entre-un-systeme-dia-et-un-modele-dia/57625>, consulté le 16 juin 2025.

Apprentissage automatique (*Machine Learning*)

Il s'agit d'un domaine de l'IA qui regroupe des techniques, méthodes et algorithmes permettant aux machines d'apprendre par elles-mêmes à partir de données sans être explicitement programmées³⁰. Si le modèle continue son apprentissage lorsqu'il est déployé en production, on peut alors parler d'apprentissage continu³¹.

Cette discipline de l'intelligence artificielle se divise en plusieurs types d'apprentissages, dont l'apprentissage supervisé, l'apprentissage auto-supervisé, l'apprentissage non-supervisé ou encore l'apprentissage par renforcement³².

L'apprentissage supervisé permet à la machine d'apprendre sur base d'un jeu de données, auxquelles on aura associé des étiquettes, des annotations (aussi appelées vérités de terrain³³), fournissant à la machine la réponse attendue. Puisque l'IA fonctionne, de façon simplifiée, sur base de probabilités³⁴, l'algorithme apprend en réduisant l'écart, l'erreur, entre sa prédiction et la réponse fournie dans l'annotation³⁵.

L'apprentissage auto-supervisé permet au système de générer lui-même des données labellisées sur base d'un corpus de données brutes³⁶.

Dans l'apprentissage non-supervisé, les données d'entraînement ne sont pas annotées. La machine apprend par elle-même en repérant des similitudes dans le jeu de données³⁷.

L'apprentissage par renforcement est un modèle apprenant par essais et erreurs, récompensé pour les bonnes décisions et améliorant ainsi la prise de décision finale.³⁸

³⁰ « Apprentissage automatique », sur *CNIL*, <https://www.cnil.fr/fr/definition/apprentissage-automatique>, consulté le 13 juin 2025.

³¹ « Apprentissage automatique » sur DataFranca, *op.cit.*, https://datafranca.org/wiki/Apprentissage_automatique, consulté le 13 juin 2025 ; « Apprentissage continu », sur *Commission Nationale de l'Informatique et des Libertés (CNIL)*, <https://www.cnil.fr/fr/definition/apprentissage-continu>, consulté le 13 juin 2025.

³² « Apprentissage automatique », *ibid.*, https://datafranca.org/wiki/Apprentissage_automatique, consulté le 13 juin 2025.

³³ « Annotations (IA) », sur *CNIL*, <https://www.cnil.fr/fr/definition/annotation-ia>, consulté le 13 juin 2025.

³⁴ Voir 1.2. Comment fonctionne l'IA.

³⁵ « Apprentissage supervisé », sur DataFranca, *op.cit.*, https://datafranca.org/wiki/Apprentissage_automatique, consulté le 13 juin 2025 ; « Apprentissage supervisé », sur *CNIL*, <https://www.cnil.fr/fr/definition/apprentissage-supervise>, consulté le 13 juin 2025.

³⁶ « Apprentissage auto-supervisé », *ibid.*, <https://www.cnil.fr/fr/definition/apprentissage-auto-supervise>, consulté le 13 juin 2025

³⁷ « Apprentissage non-supervisé », sur DataFranca, *op.cit.*, https://datafranca.org/wiki/Apprentissage_automatique, consulté le 13 juin 2025 ; « Apprentissage non-supervisé », sur *CNIL*, <https://www.cnil.fr/fr/definition/apprentissage-non-supervise>, consulté le 13 juin 2025.

³⁸ « Apprentissage par renforcement », sur DataFranca, *op.cit.*, https://datafranca.org/wiki/Apprentissage_automatique, consulté le 13 juin 2025 ; « Apprentissage par renforcement », sur *CNIL*, <https://www.cnil.fr/fr/definition/apprentissage-par-renforcement>, consulté le 13 juin 2025.

Apprentissage profond (*Deep Learning*)

L'apprentissage profond est un sous-domaine de l'apprentissage automatique. Il utilise des réseaux de neurones artificiels, agencés en plusieurs couches de neurones, chacune « traitant l'information de la couche précédente pour la transmettre à la couche suivante, ce qui permet une extraction plus fine (...) »³⁹. L'apprentissage profond se divise également en plusieurs types d'apprentissages, dont l'apprentissage supervisé, l'apprentissage non-supervisé ou encore l'apprentissage par renforcement⁴⁰.

Les réseaux de neurones artificiels « permettent la résolution de problèmes complexes, tels que la vision par ordinateur ou le traitement du langage naturel »⁴¹. Pour ce faire, il existe des architectures de réseaux de neurones spécialisées, comme les réseaux de neurones convolutifs, utilisés pour l'analyse d'images, ou les réseaux de neurones récurrents, utilisés pour l'analyse de séquences de texte ou audio. Les réseaux auto-attentionnels, appelés *transformers*⁴², utilisent le mécanisme dit « d'attention » et permettent de travailler sur des jeux de données très volumineux⁴³.

Modèle d'apprentissage⁴⁴

Il s'agit d'une structure mathématique utilisée par les algorithmes pour s'entraîner à partir des données⁴⁵. On pourrait comparer le modèle d'apprentissage à un plat cuisiné, à partir d'une recette (algorithme) appliquée à des ingrédients (données), plat qui sera ajusté et corrigé lors de la phase d'entraînement. Cependant, cette définition et comparaison du terme « modèle » est quelque peu « fourre-tout » et démontre un abus de langage, confondant le modèle, l'algorithme d'apprentissage, le réseau de neurones et son architecture. Pour le définir plus exactement⁴⁶, un modèle est la conjonction d'un ensemble d'attributs, à savoir d'un algorithme d'inférence, d'un ensemble de poids appris, ainsi que d'une topologie neuronale. Avant de définir les termes précités, il est important de comprendre la notion d'algorithme d'entraînement. Un algorithme

³⁹ « Apprentissage profond : la mécanique de la magie », sur *ISO*, <https://www.iso.org/fr/intelligence-artificielle/apprentissage-profond-deep-learning>, consulté le 13 juin 2025.

⁴⁰ « Apprentissage profond », sur DataFranca, *op.cit.*, https://datafranca.org/wiki/Apprentissage_profond, consulté le 13 juin 2025 ; « Apprentissage profond », sur *CNIL*, <https://www.cnil.fr/fr/definition/apprentissage-profond>, consulté le 13 juin 2025.

⁴¹ « Réseau de neurones artificiels (artificial neural network) », *ibid.*, <https://www.cnil.fr/fr/definition/reseau-de-neurones-artificiels-artificial-neural-network>, consulté le 13 juin 2025.

⁴² Voir 1.2. Comment fonctionne l'IA.

⁴³ « Réseau autoattentionnel », sur DataFranca, *op.cit.*, https://datafranca.org/wiki/R%C3%A9seau_autoattentionnel, consulté le 13 juin 2025.

⁴⁴ Merci à Xavier Gillard pour son explication et son aide dans la réalisation de cette définition.

⁴⁵ « Qu'est-ce qu'un modèle IA ? », sur *IBM*, <https://www.ibm.com/fr-fr/think/topics/ai-model>, consulté le 28 mai 2025.

⁴⁶ On ne considérera ici que les modèles de *deep learning*.

d'entraînement est utilisé pour permettre l'apprentissage à proprement parler. Les approches basées sur le *deep learning* utilisent un algorithme appelé « *stochastic gradient descent* » qui consiste à réaliser une inférence, c'est-à-dire à comparer la prédiction obtenue pendant l'entraînement avec la réponse attendue. Elles utilisent un critère pour « mesurer l'erreur » de prédiction, puis adaptent les poids du modèle (cf. infra) de manière à réduire cette erreur. C'est cette réduction d'erreur qu'on nomme apprentissage.

L'algorithme d'inférence est l'algorithme utilisé afin de réaliser une prédiction.

La topologie du modèle, ou architecture, est la façon dont les différentes couches neuronales sont organisées entre elles. Bien que l'algorithme utilisé soit relativement simple et le même pour tous les réseaux de neurones, il n'y a pas de limites à la manière dont ces réseaux peuvent être organisés.

Puisqu'on peut utiliser la même topologie, et le même algorithme d'apprentissage pour apprendre à une machine à réaliser des tâches qui sont fort différentes, c'est que ces éléments ne permettent pas à eux seuls de déterminer le comportement d'un modèle. En effet, les réseaux de neurones consistent en une série de "couches" qui permettent la transformation d'un signal d'entrée en un signal de sortie (appelé prédiction). Ce sont donc les poids, les paramètres appris, qui permettent de "matérialiser" les connaissances apprises par une topologie donnée.

Finetuning

Le *finetuning* est le « processus consistant à affiner un modèle préexistant sur un ensemble de données spécifiques afin de l'adapter à une tâche particulière »⁴⁷.

Chatbot

L'agent conversationnel, ou *chatbot*, est une interface qui permet de simuler un dialogue menant à la réalisation de tâches, à la réponse à une question⁴⁸.

RAG (*Retrieval-Augmented Generation*)

La génération augmentée d'information applicative (GAIA, RAG en anglais) est une approche qui combine plusieurs méthodes et qui consiste essentiellement en deux phases : le *retrieval* et la génération. Le *retrieval* est une phase de recherche asymétrique durant laquelle un modèle d'IA cherche de longs morceaux de texte pouvant répondre à une question courte. La génération est une phase durant laquelle un modèle (de type LLM, voir infra) fournit une réponse cohérente à une question posée par un utilisateur, en lui fournissant les morceaux de textes retrouvés lors

⁴⁷ « Finetuning », dans POIRIER Charles-Antoine, *AI. Termes et concepts*, Editions Futur Path, 2025, p. 4.

⁴⁸ Robot conversationnel », sur DataFranca, *op.cit.*, https://datafranca.org/wiki/Robot_conversationnel, consulté le 13 juin 2025.

de la phase de *retrieval*. Cette approche permet de limiter le risque d'hallucination en l'incitant à utiliser des informations factuellement correctes qui provenant de la base de connaissance qu'on souhaite interroger⁴⁹.

NER (*Named-entity recognition*)

La reconnaissance d'entités nommées est une « sous-tâche d'extraction d'information qui cherche à localiser et classifier les mentions d'entité nommées dans du texte non structuré en catégories prédéfinies »⁵⁰.

NLP (*Naturel Language Processing*)

Le traitement automatique du langage naturel (TALN ou TAL, NLP en anglais) est un domaine de recherche qui combine les disciplines de la linguistique, de l'informatique et de l'intelligence artificielle (*machine learning*). Il permet aux machines de comprendre, d'analyser et de générer du langage humain⁵¹. Depuis 2022 et l'avènement de ChatGPT, lorsqu'il est fait mention d'IA dans les publications grand public, il s'agit presque toujours de NLP, au point qu'aujourd'hui, le terme IA est essentiellement compris comme étant un synonyme de NLP.

LLM (*Large Language Model*)

Les grands modèles de langage (LLM en anglais), comme ChatGPT, sont les outils qui appliquent les méthodes des NLP. Ces modèles sont entraînés sur des corpus de données textuelles volumineux. Ils permettent de traduire, rédiger, générer du contenu textuel⁵².

OCR (*Optical Character Recognition*)

La reconnaissance optique de caractère (ROC ou océrisation, OCR en anglais) permet de reconnaître des caractères imprimés dans une image afin d'en extraire le texte brut, facilement exploitable en tant que texte dans un programme informatique⁵³.

⁴⁹ « Génération augmentée d'information applicative (GAIA) », *ibid.*, [https://datafranca.org/wiki/G%C3%A9n%C3%A9ration_augment%C3%A9e_d%27information_applicative_\(GAIA\)](https://datafranca.org/wiki/G%C3%A9n%C3%A9ration_augment%C3%A9e_d%27information_applicative_(GAIA)), consulté le 13 juin 2025.

⁵⁰ « Reconnaissance d'entités nommées », sur *CNIL*, https://www.cnil.fr/fr/definition/reconnaissance_dentites_nommees, consulté le 13 juin 2025.

⁵¹ « Les secrets du langage naturel décryptés », sur *ISO*, <https://www.iso.org/fr/intelligence-artificielle/traitement-langage-naturel>, consulté le 13 juin 2025.

⁵² « LLM (Large Language Model) », dans POIRIER Charles-Antoine, *op.cit.*, p. 5.

⁵³ « Reconnaissance optique de caractères », sur *DataFranca*, *op.cit.*, https://datafranca.org/wiki/Reconnaissance_optique_de_caractères, consulté le 13 juin 2025.

HTR (*Handwritten Text Recognition*)⁵⁴

La reconnaissance de texte manuscrit permet de reconnaître des caractères manuscrits dans une image afin d'en extraire le texte brut, facilement exploitable en tant que texte dans un programme informatique⁵⁵.

Hallucination

Se dit quand l'IA, généralement un *chatbot* génère une réponse inventée⁵⁶. Une hallucination de LLM se produit lorsque, en fournissant une réponse afin de compléter un prompt donné, le LLM génère un texte affirmant certains éléments sans pour autant que ceux-ci aient une quelconque matérialité dans le monde réel. Les hallucinations sont inévitables lorsqu'on travaille avec des LLM. En effet, ceux-ci ne raisonnent pas à proprement parler, et sont même incapables de comprendre le texte qui leur est soumis ou la réponse qu'ils produisent. Les LLM sont des machines à compléter du texte. Plus les modèles de langage sont avancés, plus le texte qu'ils produisent sera cohérent et vraisemblable pour un être humain, mais cela ne signifie pas pour autant que la réponse produite soit basée sur des faits réels.

Boite noire

Se dit d'un algorithme dont on ne connaît pas le fonctionnement, alors que les données d'entrées et de sorties sont connues.

XAI (*Explainable Artificial Intelligence*)

L'intelligence artificielle explicable est un domaine de recherche et un ensemble de méthodes de l'IA responsable visant à rendre les décisions des systèmes d'IA et leur production compréhensibles et interprétables par les humains⁵⁷.

2. Comment fonctionne l'IA ?⁵⁸

Comment d'une simple phrase, appelée prompt, tapée dans une boîte de dialogue, l'IA générative parvient-elle à développer une réponse structurée, compréhensible et dans un

⁵⁴ Voir 1.2.3. Traitement d'images.

⁵⁵ « Intelligence artificielle et collections de la BnF : l'exemple de l'HTR (Handwritten Text Recognition) », sur *Bibliothèque nationale de France*, <https://www.bnf.fr/fr/intelligence-artificielle-et-collections-de-la-bnf-lexemple-de-lhtr-handwritten-text-recognition>, consulté le 13 juin 2025.

⁵⁶ « AI Key Terms », sur *Chapman University*, <https://www.chapman.edu/ai/key-terms.aspx>, consulté le 13 juin 2025.

⁵⁷ « Explicabilité », sur *DataFranca*, *op.cit.*, <https://datafranca.org/wiki/Explicabilit%C3%A9>, consulté le 13 juin 2025 ; « Explicabilité (IA) », sur *Commission Nationale de l'Informatique et des Libertés (CNIL)*, <https://www.cnil.fr/fr/definition/explicabilite-ia>, consulté le 13 juin 2025.

⁵⁸ Cette section a été rédigée par Camille Berny, avec l'aide et la relecture attentive de Xavier Gillard, et sur base des articles « ChatGPT, un beau parleur bien entraîné », sur *LINC, Laboratoire d'Innovation Numérique de la*

langage irréprochable ? Comment comprend-elle la question que nous lui posons ? Comment peut-elle générer des métadonnées à partir d'une pièce comptable ? Comment parvient-elle à indexer des comptes-rendus ? Comment parvient-elle à décrire une image ou à déchiffrer de la paléographie à partir du scan d'une charte médiévale ?

Je vous invite à faire un tour sous le capot, afin de découvrir, en pratique et en technique, comment l'IA fonctionne. Et puisqu'il n'y a pas qu'une seule IA, nous étudierons le fonctionnement des modèles génératifs, et plus particulièrement celui des LLM (*Large Language Models*)⁵⁹, à partir du prompt « quel est l'impact de l'IA sur la pratique archivistique ? » et des modèles discriminatifs (non-génératifs).

a. Modèle génératif

Un modèle d'IA génératif permet notamment de générer du langage naturel, de répondre à des questions, de traduire, de résumer des documents. Certains sont spécialisés⁶⁰ dans la génération de texte⁶¹ (comme ChatGPT), mais d'autres peuvent produire des images (comme Dall-E), des sons, ou des vidéos. Les modèles peuvent également collaborer entre eux.

Grâce à l'IA générative, l'archiviste peut résumer des documents, les traduire, les décrire ou encore en extraire les métadonnées.

CNIL, <https://linc.cnil.fr/dossier-ia-generative-chatgpt-un-beau-parleur-bien-entraine>, consulté le 11 juin 2025 ; ALAMMAR Jay, « The illustrated Word2vec », <https://jalammar.github.io/illustrated-word2vec/>, consulté le 17 juin 2025 ; « Comment une IA générative crée-t-elle du texte ? », sur *Délégation Régionale Académique Numérique Educatif*, <https://drane-versailles.region-academie-idf.fr/spip.php?article845>, consulté le 11 juin 2025 ; GRINBAUM Alexei e.a., « Systèmes d'intelligence artificielle générative : enjeux d'éthique ». Comité national pilote d'éthique du numérique, 2023, <https://cea.hal.science/cea-04153216/>, consulté le 6 juin 2025, LORRE Benjamin, *L'intelligence artificielle générative. Explications, modèles, enjeux*, Paris, ISTE Editions, 2025 ; MITTAL Aayush, « NLP Rise with Transformer Models | A Comprehensive Analysis of T5, BERT, and GPT », sur *Unite.ai*, <https://www.unite.ai/nlp-rise-with-transformer-models-a-comprehensive-analysis-of-t5-bert-and-gpt/>, consulté le 11 juin 2025 ; MUEHLBERGER Guenter e.a., « Transforming scholarship on the archives through handwritten text recognition. Transkribus as a case study », dans *Journal of Documentation*, vol 75, n°5, 2029, p. 954-976 ; AZAD Reza e.a., « Advances in medical image analysis with Transformers : a comprehensive review », dans *Medical Image Analysis*, vol 91, 2024 ; WU Xingjiao, « A survey of human-in-the-loop for machine learning », dans *Future Generation Computer System*, vol. 135, 2022, p. 363-381 ; CNIL, <https://www.cnil.fr/fr/intelligence-artificielle/glossaire-ia>.

L'outil Copilot a été utilisé pour proposer l'exemple de tokenisation et de vectorisation, à partir de la question « Dans mon mémoire de master sur l'influence de l'IA sur la pratique archivistique, j'écris un chapitre sur le fonctionnement technique de l'IA. Je souhaite y parler précisément des vecteurs, *transformers*, etc. afin d'être la plus complète et précise possible, en vulgarisant afin que mon audience comprenne le fonctionnement de l'IA. Si je tape un prompt, que se passe-t-il réellement pour que l'IA parvienne à me produire une réponse. J'aimerais donner un exemple de vecteur et de *transformer*, un exemple dans le domaine de l'archivistique. Peux-tu le donner le vecteur du mot "archives" dans la phrase "quel est l'impact de l'IA sur la pratique archivistique" ».

⁵⁹ *Large Language Models*, qui pour rappel, sont des décodeurs générant du texte.

Voir 1.1. Définitions de base.

⁶⁰ IA génératives monomodales.

⁶¹ Modèle conçu pour comprendre et générer du langage naturel, il s'agit des modèles GPT (*Generative Pre-trained Transformer*).

Tokenisation

Lorsque, via notre interface utilisateur, nous introduisons des phrases, nous posons des questions, l'IA ne les comprend pas. Elle n'est pas capable de comprendre ce que nous lui demandons. La première phase du processus est celle de la tokenisation : le texte encodé est découpé en unités, appelées *tokens*. Il s'agit de mots, de parties de mot, de ponctuation, etc.

Les différents *tokens* de notre prompt seront : "Quel", "est", "l'", "''", "impact", "de", "l'", "''", "IA", "sur", "la", "pratique", "archivistique", "?".

Vectorisation

L'IA ne comprenant pas les mots ni leur sens, il est nécessaire d'utiliser un langage compréhensible par la machine. Chaque *token* est donc transformé en vecteur numérique (appelé *embedding*), une série de nombres qui définissent des informations sur le sens du mot, son usage, son contexte, ses relations avec d'autres mots, etc. Les *embeddings* utilisent en quelque sorte les mathématiques pour capturer les relations sémantiques entre les différents mots d'une phrase.

Le mot « archivistique » par exemple, deviendra un vecteur tel que [0.12, -0.45, 0.33, ..., 0.08], mais avec plusieurs centaines de dimensions. Ces vecteurs ont été appris à partir de milliards de textes lors de la phase d'apprentissage. Ils permettent à l'IA de lier le mot « archivistique » à ceux d'« archives », de « conservation », de « patrimoine », etc.

Transformers

Le modèle d'IA génératif utilise une architecture de réseaux de neurones, afin d'utiliser les *embeddings* et de les transformer en vecteurs enrichis qui prennent en compte le contexte de la phrase et identifient les relations entre les mots. Au sein d'un réseau de neurones, nous retrouvons des blocs, appelés *transformers*⁶², qui permettent d'analyser les *embeddings* sous différents points de vue et de raffiner leur contextualisation.

C'est le mécanisme d'attention qui, pour chaque mot donné, pondère l'importance des autres mots de la phrase par rapport à ce mot. Il permet de faire le lien entre les mots pertinents. Ces opérations sont répétées plusieurs fois afin d'affiner le résultat. Dans notre exemple,

⁶² Composés d'un sous-module de self-attention multi-tête, d'une couche de normalisation et d'une couche *feed forward*. La couche de self-attention multi-tête permet à chaque tête de ce module d'adopter un point de vue pour analyser les *embeddings* sous différents aspects, afin d'amener plus de précision. La couche de normalisation est un artefact technique permettant au modèle de bien fonctionner, et cela même dans des situations qui diffèrent de son modèle d'apprentissage. Enfin, la couche *feed forward* est une couche neutrale qui combine les résultats des différents modules de self-attention en un seul résultat, raffinant ainsi la contextualisation des *embeddings*.

« impact » est le mot central, « IA » en est le sujet, « pratique archivistique » en est le domaine d'application.

En pratique, les architectures basées sur les *transformers* sont de trois types, utilisés selon différents cas : les modèles « *encoders and decoders* » sont principalement utiles pour de la traduction automatique, les modèles « *decoders only* » sont utilisés pour des tâches génératives et les modèles « *encoders only* » interviennent principalement dans les tâches de classification et de NER (*Named-entity Recognition*)⁶³.

Le *transformer* comprend donc plusieurs couches, au sein desquelles les vecteurs sont enrichis pour y apporter plus de contexte et « prédire » éventuellement⁶⁴ le mot suivant, en fonction de ce que le modèle a appris durant son apprentissage. Ces vecteurs finaux permettent d'établir des probabilités sur le mot le plus pertinent qui pourrait suivre le mot « impact ». Ainsi, lors de son entraînement, le modèle a appris des motifs dans des livres, articles scientifiques, sites internet, dialogues, etc. Il peut donc prédire que le mot « impact », qui est fortement lié aux mots « IA » et « pratique archivistique », sera probablement suivi de mots tels que « classement », « indexation », « recherche ». L'IA donne un score à chacun de ces mots, celui qui obtient le score le plus haut est le plus probable et sera donc choisi par l'IA pour compléter la phrase et répondre à la question. La réponse la plus probable à la question de départ pourrait donc être « L'intelligence artificielle peut transformer la pratique archivistique en automatisant le classement, l'indexation et la recherche de documents ».

b. Modèle discriminatif

Les modèles discriminatifs sont également des modèles *transformers*, a priori composés uniquement d'encodeurs. Ils ne peuvent donc pas générer de texte. Requéran un seul passage dans le modèle (pour fournir une seule prédiction⁶⁵), ils sont souvent moins couteux en ressources et plus précis pour des tâches spécifiques, et permettent à l'archiviste de classer, d'indexer des documents. Ce sont des modèles de traitement de langage naturel non-génératif⁶⁶, conçus pour comprendre le sens d'un texte.

Les étapes de tokenisation (en petites parties de mots) et de vecteurisation (pour saisir le sens du mot dans son contexte) sont similaires à celles précédemment présentées. Dans l'étape *transformer*, le modèle discriminatif peut lire tous les mots de la phrase donnée dans le

⁶³ Voir 2.2. Modèles discriminatifs.

⁶⁴ Les modèles *encoder only* ne font pas que de la « prédiction » de mots. Ils peuvent aussi calculer des *embeddings* qui seront stockés dans une base de données, ou encore évaluer un sentiment exprimé dans un texte.

⁶⁵ Contrairement aux modèles génératifs qui nécessitent plusieurs passages à travers le modèle pour prédire à chaque fois un *token*.

⁶⁶ Comme le modèle BERT (*Bidirectional Encoder Representations from Transformers*).

prompt, il comprend donc le sens global de la phrase. Ces modèles produisent des vecteurs contextualisés pour chaque *token*. Ces vecteurs peuvent ensuite être utilisés pour classer un texte, trouver des réponses dans un document, extraire des noms, des lieux, des dates, etc. Ils sont spécialisés dans la découverte d'un mot masqué, un mot manquant dans une phrase et dans un contexte spécifique, et pas dans la prédiction de mots pour générer des phrases automatiquement, comme dans les modèles génératifs.

Ces modèles peuvent être entraînés sur un corpus d'archives défini pour reconnaître des types de documents, une période historique en particulier, ou encore des niveaux de confidentialité.

Notons que le RAG⁶⁷ est un mélange entre un modèle discriminatif (par la recherche d'informations dans des documents pertinents dans un corpus donné) et de modèle génératif (par la production d'une réponse contextualisée).

c. Traitement d'images

Qu'en est-il des images traitées par l'IA ? L'archiviste peut être confronté à deux cas de figure principaux : l'un permettra de reconnaître du texte paléographié, il s'agit de la tâche HTR, l'autre pourra aider l'archiviste à classer ou indexer les images selon leur contenu.

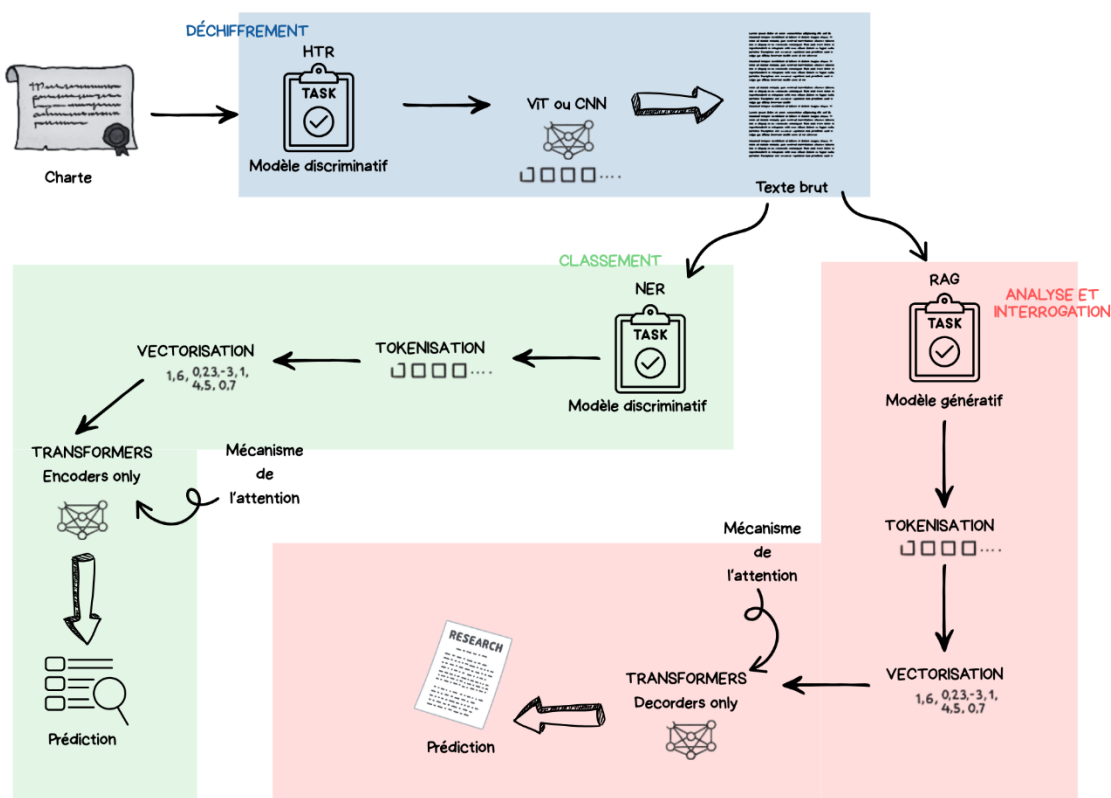
Le premier usage, la tâche HTR, est réalisé grâce à des réseaux de neurones convolutifs (CNN) ou à des *Visions Transformers* (ViT). Les CNN consistent en une série de couches, fonctionnant comme des filtres, qui vont reconnaître des motifs au sein de l'image, tandis que les ViT permettent de découper l'image en petites « tuiles », traitées comme des *tokens*. Certains modèles multimodaux (comme Gemini) peuvent traiter à la fois image et texte, produisant des *embeddings* proches pour une image et un mot similaires.

Un modèle génératif d'IA peut aider au classement et à l'indexation d'une image. L'image est transformée en vecteurs permettant d'identifier son contenu sémantique. Les vecteurs sont traités par des *transformers* spécifiques aux images, qui permettent de comparer ce vecteur produit à des vecteurs produits à partir de texte. Par exemple, une image d'archives représentant le Roi Albert 1^{er} sur le front en 1917 aurait un vecteur très similaire au vecteur du texte « Le Roi Albert 1^{er} inspecte les troupes belges sur le front en 1917 ». Sur base de la comparaison de ces vecteurs et de probabilités, la description générée pourrait être choisie pour décrire la photo ou l'indexer.

⁶⁷ Voir 1. Définitions de base.

3. Conclusion

À titre de conclusion, et afin de rassembler les différentes notions développées précédemment, nous souhaitons nous mettre dans la peau d'une archiviste qui doit analyser une charte du 14^e siècle à l'aide de l'IA : nous souhaiterions déchiffrer le document, mais aussi l'interroger en vue de produire un article, une notice ou une description et les classer dans une base de données. En résumé et en schéma, voici comment cela se passe⁶⁸ :



⁶⁸ Schéma réalisé par Camille Berny.

Première partie. La pratique archivistique face aux enjeux du numérique avancé.

Si le numérique s'est progressivement développé à la fin du 20^e siècle pour exploser dans les années 2000, son utilisation dans les institutions et les entreprises a été le reflet de son évolution, à tel point que le papier s'est vu être en partie remplacé par des fichiers virtuels, entraînant un chamboulement chez les archivistes. Données, documents et informations se confondent et se multiplient. De nouveaux formats et supports apparaissent, entraînant avec eux des défis pour la pratique archivistique. Les archives, qu'elles soient courantes, intermédiaires ou définitives, se mélangent dans les ordinateurs ; les lieux de stockages et de conservation traditionnels sont revus.

Avant d'aborder l'impact de l'IA sur la pratique archivistique tout au long du cycle de vie des archives, il s'agit de voir comment le numérique l'a modifiée dans un premier temps. Comment assurer la conservation à long terme de ces fichiers numériques, leur pérennité, dont on sait que les formats ne sont pas garantis dans le temps ? Comment garantir leur fixité et leur intégrité, alors qu'ils peuvent être copiés et modifiés à souhait ? Comment redéfinir les archives ? Comment le métier a-t-il évolué ? Comment les fondements de l'archivistique, dont le principe du respect des fonds et le cycle de vie des documents, ont-ils été reconfigurés par l'arrivée du numérique ? Il ne s'agira pas ici de proposer une étude exhaustive des transformations du numérique en archivistique, mais de se concentrer sur les aspects qui apparaissent aujourd'hui les plus directement réinterrogés par l'intelligence artificielle. Ce chapitre proposera ainsi un tour d'horizon des principales mutations documentaires, techniques et professionnelles identifiées dans la littérature comme préparant ou accompagnant l'intégration de l'IA dans les pratiques archivistiques.

Cette partie lira donc l'effet du numérique sur la pratique archivistique par paliers. Si l'idée première était de suivre les différentes étapes des trois âges, cette structure est apparue scolaire. Au fil de nos lectures, il est par contre apparu logique de voir en quoi il reconfigure la forme documentaire des archives au sens large et les conséquences sur leur authenticité et leur intégrité dans un premier temps, puis comment il transforme les dispositifs de conservation avant d'entraîner un déplacement des responsabilités professionnelles. C'est seulement dans le prolongement de ces mutations que la question du cycle de vie se pose, comme un cadre mis à l'épreuve par de nouvelles temporalités documentaires.

1. Le numérique et la forme documentaire. Redéfinir l'authenticité, l'intégrité et le lien archivistique à l'ère numérique

« Le numérique ne change pas le stockage de documents inutiles, périmés, inexploitable, car cela relève plus de mauvaises pratiques humaines que de la technique et de la technologie ; au contraire, le numérique ne fait qu'exacerber ces mauvaises pratiques et provoque des entassements bien pires que dans les années 1960 »⁶⁹.

L'une des premières observations que l'on peut formuler avec l'arrivée du numérique, c'est le volume de documentation produit qui a explosé, de même que sa « dissémination »⁷⁰. Cette « volumétrie » peut être évidemment caractérisée par une quantité exponentielle du nombre de fichiers, mais également par une grande variété de supports différents, de producteurs et de lieux de production multiples⁷¹. Des fichiers intermédiaires sont automatiquement enregistrés, des copies sont créées et stockées dans différents dossiers, entraînant un vrac numérique important. La production de documents numériques, en plus d'être volumineuse et dispersée dans les formats et supports, a également entraîné une « certaine forme d'impensé des questions liées à la fiabilité, à l'exactitude et à l'authenticité »⁷².

En effet, les copies et fichiers intermédiaires, de même que le fait que contenu et support soient dissociables, amènent la question de la fixité des documents et un premier enjeu de l'archivage numérique : « le support n'est plus en capacité de garantir la fixité du document d'archives et les objets nativement numériques ont justement comme qualité intrinsèque d'être instables »⁷³. Puisque les fichiers peuvent être modifiés, mis à jour, le contenu informationnel du document et sa forme s'en trouvent transformés, sans que cela puisse être visible sur le support (les traces des modifications sont enregistrées, mais de façon invisible). Le propos archivistique s'est ainsi progressivement déplacé de la fixité des documents vers les notions d'intégrité et d'authenticité. C'est désormais sur ces dimensions que les professionnels et la littérature, parmi lesquels le projet *InterPARES*, insistent particulièrement. L'intégrité est définie comme « *the quality of being complete and unaltered in all essential aspects. With*

⁶⁹ CHABIN Marie-Anne, « Les deux âges des archives », sur *Le blog de Marie-Anne Chabin*, novembre 2022, <https://www.marieannechabin.fr/2022/11/les-deux-ages-des-archives/>, consulté le 19 juin 2025.

⁷⁰ ID., « L'archiviste de 2030 entre archives numériques et utilisateurs connectées », dans SERVAIS Paul et MIRGUET Françoise (éd.), dans SERVAIS Paul et MIRGUET Françoise (éd.), *Archivistes de 2030. Réflexions prospectives*, Louvain-la-Neuve, L'Harmattan, 2015, p.63.

⁷¹ *Ibid.*, p.63.

⁷² GUYON Céline, « L'archivage comme dispositif de transformation de la nature intrinsèque des objets nativement numériques », dans *Balisages*, n°1, 2020, <http://journals.openedition.org/balisages/282>.

⁷³ *Ibid.*

identity, a component of authenticity »⁷⁴, tandis que l'authenticité est « *the trustworthiness of a record as a record; i.e., the quality of a record that is what it purports to be and that is free from tampering or corruption* »⁷⁵. Les éléments essentiels dont il est question pour maintenir l'intégrité sont la structure et le contenu (liés à l'organisation du contenu), l'identifiant et le contexte (donnés par des métadonnées) et certains éléments de forme. Les supports réinscriptibles, autorisant les modifications des fichiers numériques, ne permettent quant à eux plus de garantir l'intégrité et l'authenticité des documents, ils deviennent dès lors des « éléments de contexte technologique »⁷⁶.

Pour pallier ces risques, différentes solutions sont proposées, comme par exemple le transfert des documents, dès leur création, vers un système d'archivage électronique des documents, ou encore une « conservation contextualisée »⁷⁷ qui permet d'enregistrer les modifications, datées, effectuées dans un document. Ces solutions sont gages d'une variabilité limitée et de traçabilité pour les documents numériques, qui peuvent dès lors être considérés comme des archives dignes de confiance⁷⁸, pour autant qu'une vision archivistique ait été associée aux documents dès leur conception.

Un second enjeu répertorié dans la littérature est celui du principe du respect des fonds. Dans le monde analogique, il est normalement assez facile de comprendre l'organisation des documents entre eux, leur classement par le producteur. On peut repérer la place de chacun par rapport aux autres dans un dossier, mais aussi par rapport à une série ou à un fonds d'archives : le lien archivistique est en principe lisible. Avec le numérique, même si les différents fichiers sont enregistrés dans des dossiers, sous-dossiers, etc., et répondent à une logique de classement personnelle ou institutionnelle (avec ses règles et procédures internes), « il faut cependant convenir que ces liens entre le document, le dossier, le processus, l'activité et la fonction nous apparaissent comme essentiellement virtuels »⁷⁹. De plus, le recours croissant aux moteurs de recherche pour accéder aux documents tend à reléguer les arborescences documentaires au second plan. Le fonds d'archives dans son ensemble n'est plus facilement lisible.

⁷⁴ « Integrity », sur *InterPARES Trust*, <http://interparestrust.org/terminology/term/integrity>, consulté le 1 juillet 2025.

⁷⁵ « Authenticity », *ibid.*, consulté le 1 juillet 2025.

⁷⁶ DURANTI Luciana et THIBODEAU Ken, « The Concept of Record in Interactive Experiential and Dynamic Environments: the View of InterPARES », dans *Archives and Museum Informatics*, vol.6, n°1, mars 2006, p.18.

⁷⁷ GUYON Céline, « L'archivage comme dispositif de transformation de la nature intrinsèque des objets nativement numériques », *op.cit.*.

⁷⁸ *Ibid.*

⁷⁹ NAHUET Robert, « L'archivistique contemporaine à l'âge adulte : pertinence et actualité du respect des fonds », dans *Archives*, vol.41, n°1, 2009-2010, p.53.

Afin de maintenir le lien archivistique entre les documents numériques, les métadonnées sont préconisées. Elles permettent d'identifier les fichiers, mais aussi de rattacher un document à l'ensemble auquel il appartient, ce qui est une condition nécessaire au maintien de l'intégrité du document numérique⁸⁰. De plus, les métadonnées permettent de garantir « le contexte de création des documents »⁸¹, « d'assurer que les documents constituent une preuve de la réalisation d'une activité, donc de leur qualité de témoin irréfutable »⁸².

2. Le numérique et les dispositifs de gestion et de conservation. Garantir la valeur probante et la mémoire à long terme des archives numériques

Préalablement à l'arrivée dans le dépôt ou le centre d'archives, les archives définitives en devenir doivent passer par une étape d'évaluation. Si l'évaluation intellectuelle, qui permet d'établir que « leur provenance, leur ordre, leur complétude, leur authenticité, leur fiabilité et aussi leur valeur intrinsèque »⁸³ reste identique, une évaluation technique des documents numériques est également nécessaire afin de déterminer si le centre d'archives peut assurer la maintenance des formats et des supports afin de garantir une conservation pérenne et un accès aux archives sur le temps long.

Comme dans toutes les institutions, les dépôts et centres d'archives sont également passés au numérique dans les années 2000. Les procédures ont été informatisées, amenant la « mise en œuvre de gestion de flux métier »⁸⁴ : depuis les bordereaux de versement qui sont complétés dans des logiciels, jusqu'aux répertoires d'archives, registres ou même fonds d'archives entiers qui ont ainsi été numérisés, en passant par la gestion des éliminations dans les systèmes d'information⁸⁵. De même, le langage XML est utilisé à travers les formats DTD EAD et DTD EAC afin de structurer les descriptions archivistiques numériques et de garantir leur conformité aux normes ISAD(g) et ISAAR (CPF), tout en permettant « l'échange de notices structurées entre plusieurs partenaires et portails »⁸⁶.

⁸⁰ DURANTI Luciana, « The Archival Bond », dans *Archives and Museum Informatics*, vol.11, n°3, septembre 1991, p.216.

⁸¹ NAHUET Robert, *op. cit.*, p.54.

⁸² *Ibid.*

⁸³ BASMA Makhoul Shabou, « Fonction d'évaluation des archives. Bilan sommaire des développements, des enjeux actuels et des défis futurs », dans GAGNON-ARGUIN Louise et LAJEUNESSE Marcel (éd.), *Panorama de l'archivistique contemporaine. Évolution de la discipline et de la profession. Mélanges offerts à Carol Couture*, Québec, Presses de l'Université du Québec, 2015, P.200.

⁸⁴ BANAT-BERGER Françoise, « Les fonctions de l'archivistique à l'ère du numérique », dans DELPIERRE Nicolas, HIRAUX Françoise et MIRGUET Françoise (éd.), *Les chantiers du numérique. Dématérialisation des archives et métiers de l'archiviste*, Louvain-la-Neuve, L'Harmattan, 2012, p.40.

⁸⁵ *Ibid.*

⁸⁶ *Ibid.*

Dans le monde analogique, lorsque des archives sont traitées, « le contenu de l'information dans son contexte de production » est géré, « en intégrant tous les éléments permettant de renseigner, de qualifier cette information tout au long de son cycle de vie »⁸⁷. Une description permet ensuite de détailler le contenu du document et de le rattacher à une série documentaire ou un fonds⁸⁸. Dans le monde numérique, le principe reste le même puisqu'un « objet documentaire » est composé « de l'ensemble des fichiers organisés en dossiers et sous-dossiers relatifs à ce texte », mais aussi des « différentes versions du projet jusqu'à ses formes finales exprimées sous format PDF, mais également XML », des « métadonnées décrivant ce texte », de même que l'ensemble des modifications qu'a pu subir le document⁸⁹. Peuvent également être conservé dans ces dossiers les empreintes numériques, les certificats d'horodatage, les signatures électroniques ou encore les autorisations d'accès et de modifications des fichiers⁹⁰.

Les enjeux du passage au numérique sur les archives définitives sont avant tout ceux de la pérennité, de l'intégrité, de la confidentialité et de l'accessibilité. La pérennité est primordiale pour que les documents électroniques archivés restent intègres, lisibles et accessibles. Il s'agit de mettre en place des « outils techniques et organisationnels »⁹¹ afin que les supports de stockage, les logiciels et le matériel restent à jour. De même des technologies permettent de garantir l'intégrité des documents archivés, grâce à des « dispositifs d'empreinte », des fonctions de *hashage*, à savoir des calculs algorithmiques permettant de « vérifier que le contenu d'un document est resté intact, dans le temps, toute modification du document entraînant automatiquement une empreinte différente »⁹².

Les archives courantes, qui sont des « objets numériques dynamiques »⁹³, deviennent définitives lors de l'archivage dans un système d'archivage électronique, c'est-à-dire lorsque la forme devient fixe (« au travers du dispositif logique formé par les métadonnées »⁹⁴) et que le contenu devient stable (« au travers du dispositif de traçabilité »⁹⁵).

⁸⁷ *Ibid.*

⁸⁸ *Ibid.*

⁸⁹ *Ibid.*, p.41.

⁹⁰ *Ibid.*

⁹¹ DEMOULIN Marie, « Quelques aspects juridiques de l'archivage électronique », dans DELPIERRE Nicolas, HIRAUX Françoise et MIRGUET Françoise (éd.), *op.cit.*, p.21.

⁹² GUYON Céline, « L'archivage comme dispositif de transformation de la nature intrinsèque des objets nativement numériques », *op.cit.*

⁹³ ID., « Une archivistique sous influence. La pratique archivistique face aux innovations technologiques (1960-2022) », dans *Revue d'Histoire culturelle, XVIIIe-XXIe siècles*, n°5, 2022, p.13, <https://journals.openedition.org/rhc/3466>.

⁹⁴ *Ibid.*

⁹⁵ *Ibid.*

Ainsi, la perte de la capacité du support à garantir la fixité entraîne un déplacement des garanties archivistiques, désormais fondées sur l'intégrité informationnelle, la traçabilité et les métadonnées.

3. Le numérique et l'évolution des pratiques professionnelles. De la conservation à la gouvernance informationnelle : transformations du métier d'archiviste à l'ère numérique

Pérotin et la théorie des trois âges en archivistique préconisent une intervention de l'archiviste assez tôt dans le cycle de vie. En effet, lorsque Pérotin a développé la théorie, il estimait que les archivistes devaient intervenir dès le premier âge des archives pour « obtenir que les bureaux fabriquent de bonnes archives et constituent des dossiers que n'encombrent pas les inutilités »⁹⁶. Au vu des modifications qu'amène le numérique dans la pratique archivistique, notamment pour garantir l'intégrité des documents ou le respect des fonds, grâce, entre autres, à l'ajout de métadonnées, il apparaît que les archivistes doivent intervenir au plus tôt dans le cycle de vie des archives, mais pas seulement à titre de conseil pour désencombrer les dossiers.

Avec l'avènement du numérique, la gestion documentaire ne peut plus se limiter au cycle de vie traditionnel des archives. La production massive, instantanée et souvent collaborative des documents numériques brouille les frontières entre les âges définis par Pérotin : un même fichier peut être réutilisé, modifié, partagé ou copié, et coexister en plusieurs versions suivant chacune leur propre trajectoire, rendant plus difficile l'application d'une logique linéaire des archives courantes, intermédiaires et définitives. De même, les exigences croissantes en matière de traçabilité, de conformité réglementaire et de preuve juridique imposent une gestion rigoureuse dès la création du document. Ainsi, avec le numérique, la gestion des archives tout au long de leur cycle de vie, dès leur validation, s'est développée et a mué vers un cycle de vie de l'information⁹⁷. Cette gestion est plus connue sous le nom de *records management*. En anglais, *record* est l'un des termes employés pour qualifier un document au cours du cycle de vie, au même titre que *document* et *archive* (en anglais toujours). Un *document* est « une information enregistrée sur un support »⁹⁸. Si le *document* « reste dans un état informatif sans rôle particulier en termes de preuve »⁹⁹, on le qualifiera de *non-record*.

⁹⁶ PÉROTIN Yves, *op.cit.*, p.33.

⁹⁷ KERN Gilliane, HOLGADO Sandra, COTTIN Michel, *op.cit.*, p.54

⁹⁸ CHABIN Marie-Anne, « Le Records management : concepts et usages », 2012, <https://www.marieannechabin.fr/arcateg/wp-content/uploads/2017/03/MAC-Le-Records-management.-Concept-et-usages-2012.pdf>, consulté le 17 juin 2025.

⁹⁹ *Ibid.*

Au contraire, si le *document* est validé, il deviendra un *record*. Les *archives* sont quant à elles des *records* qui sont conservés en raison de leur valeur patrimoniale, elles correspondent à la dernière étape du cycle de vie des archives.

Si la terminologie francophone du cycle de vie se veut plus relative au lieu de conservation des archives, à leur dimension logistique (dans les bureaux pour les archives courantes, dans des locaux séparés pour les archives intermédiaires, aux archives pour les archives définitives), l'usage anglophone de la terminologie archivistique met en avant et insiste sur la valeur juridique de l'information. Le *records management* se focalise sur les archives courantes et les archives intermédiaires, sur les *records* donc, les documents validés, et avant archivage. On privilégie cette notion dans un cadre numérique, puisque la dimension logistique n'a plus de sens. La norme ISO 15489 le définit comme le « champ de l'organisation et de la gestion en charge d'un contrôle efficace et systématique de la création, de la réception, de la conservation, de l'utilisation et du sort final des documents, y compris des méthodes de fixation et de préservation de la preuve et de l'information liées à la forme des documents »¹⁰⁰. On retrouve bien dans cette définition les notions de cycle de vie des archives, de même que celle de valeur probante. Selon la norme ISO 15489, il faut pouvoir garantir l'authenticité d'un *record*, sa fiabilité, son intégrité et son exploitabilité¹⁰¹. En effet, les risques pour la gestion documentaires sont importants, et davantage encore dans un environnement numérique : il est bien sûr question de pérennité, mais aussi de la modification d'un *record* (donc un document validé), entraînant une modification de sa valeur probante¹⁰², ou encore dans la perte d'un document, noyé dans la masse documentaire, et mal renseigné par les métadonnées¹⁰³.

Le *records management* s'attache donc aux archives courantes et intermédiaires, dans des projets de *records management* (ou des projets de gestion électronique des documents pour les archives courantes plus particulièrement), sans que la frontière ne soit malgré tout bien définie avec les archives définitives (qui sont conservées dans des systèmes d'archivage électronique)¹⁰⁴. Si les outils sont différents, leur gestion, dès la création d'un document, peut être confiée à un archiviste. En effet, avec le glissement de terminologie amené par le numérique

¹⁰⁰ Norme ISO 15489-1:2016, *Information et documentation – Gestion des documents d'activité*.

¹⁰¹ *Ibid.*

¹⁰² Sur cette question, le format et le contrôles des accès en écriture sont essentiels.

¹⁰³ CHABIN Marie-Anne, « Le Records management : concepts et usages », *op.cit.*

¹⁰⁴ Services des ressources informationnelles et archives de l'Université de Lausanne, « Records management et gouvernance informationnelle : Concepts et explications », 2016, p.8, https://www.unil.ch/files/live/sites/unil/files/02-universite/0204-organisation/020404-unites-et-services/UNIRIS/references/UNIL_RM_Gouvernance_Info_concepts_explications_VF-1.pdf, consulté le 28 septembre 2025.

et les risques engendrés par celui-ci, Marie-Anne Chabin positionne la nouvelle gestion archivistique sur l'ensemble du cycle de vie, en particulier sur les « métadonnées et les droits liés aux archives » (on pense notamment aux questions d'accès et d'autorisation selon le profil des utilisateurs) et leur contrôle¹⁰⁵. Le *records management* ne constitue pas un simple prolongement technique de l'archivistique, mais un déplacement du rôle même de l'archiviste vers une fonction de gouvernance de l'information¹⁰⁶

4. Le numérique et la mise à l'épreuve du cycle de vie des archives.

Le numérique et la remise en question du cycle de vie des archives

Avec le numérique se pose la question du sens de la théorie des trois âges de Pérotin, puisque les frontières logistiques qui définissent les âges des archives analogiques ne sont plus pertinentes dans un monde électronique. De même, la question de l'intégrité et de l'authenticité des documents se pose dès leur création, et pas juste au moment de leur conservation¹⁰⁷.

Pourtant, différents arguments poussent à maintenir les trois âges. D'une part, selon la Commission Nationale de l'Informatique et des Libertés (CNIL, France), cette division reste pertinente, car elle permet de définir des autorisations d'accès différents selon les stades, allant vers plus de restriction. De même cette « distinction reste également opérante concernant l'accès à l'information qui, durant les deux premiers âges, doit rester celle qui est le plus familière à l'utilisateur initial »¹⁰⁸. Enfin, « en termes de responsabilités du service producteur »¹⁰⁹, la division reste également pertinente, puisque ce n'est qu'à la fin de la durée d'utilité administrative, soit à la fin de l'âge des archives intermédiaires, que le producteur doit décider de conserver ou d'éliminer les documents.

Si d'autres chercheurs ont proposé des adaptations aux trois âges du cycle de vie, selon le statut *offline*, *nearline* ou *online* du document, ou encore selon le traitement à réserver aux

¹⁰⁵ CHABIN Marie-Anne, « L'archiviste de 2030 entre archives numériques et utilisateurs connectés », *op.cit.*, p.68.

¹⁰⁶ « La littérature scientifique en archivistique définit la gouvernance comme étant l'ensemble des lois, des règlements, des politiques, des procédures et des lignes directrices dont le déploiement vise à encadrer les pratiques de gestion de l'information documentaire (Smallwood, 2019). Autrement dit, il est question d'un ensemble d'outils d'encadrement misant sur l'harmonisation des pratiques de gestion de l'information, et ce, en conférant à chaque acteur organisationnel (ex. dirigeants, experts, agents administratifs, etc.) des rôles et des responsabilités dans la création, la gestion, la diffusion et la préservation à long terme de l'information ».

ALAOUI Siham, « L'intelligence artificielle et la gestion documentaire : quels apports ? Quels enjeux ? », dans *La Revue canadienne des sciences de l'information et de bibliothéconomie*, vol. 46, n°2, 2023, p.11.

¹⁰⁷ FUENTES-HASHIMOTO Lourdes, « L'archivage électronique à l'épreuve de la théorie des trois âges » sur *Archives Online*, <https://archivesonline.wordpress.com/2010/08/30/larchivage-electronique-a-lepreuve-de-la-theorie-des-trois-ages/>, consulté le 24 juin 2025.

¹⁰⁸ BANAT-BERGER Françoise, *op.cit.*, p.51.

¹⁰⁹ *Ibid.*

documents (capture, disposition, évaluation, préservation, transfert ou destruction)¹¹⁰, certains ont développé une tout autre théorie, dans la lignée de l'archivistique post-moderne¹¹¹ : en 1996, Frank Upward formalise et structure le *records continuum*, modèle qui a émergé en Australie dès les années 1960, afin d'allier les besoins des archivistes et des *records managers*¹¹². Selon eux, le numérique permet trop de changements et de modifications dans un document pour rester dans la linéarité des trois âges. En collaboration avec Sue McKemish et Barbara Reed, Frank Upward a la volonté de « mieux assumer la logique du continuum selon laquelle évoluent les archives dans le contexte électronique en assurant « *the continuum of responsibilities* » entre les gestionnaires de documents et les archivistes »¹¹³. Selon le principe du continuum des archives, en français, celles-ci « sont perçues comme étant « toujours en devenir », c'est-à-dire qu'elles sont toujours susceptibles d'être transformées par de nouveaux utilisateurs dans des contextes nouveaux »¹¹⁴.

Le *records continuum* est un modèle espace-temps, défini par l'*Australian Standard*, qui établit les normes en Australie, comme « un régime cohérent et intégré de processus de gestion des documents, allant de leur création (et même avant, lors de la conception des systèmes de gestion documentaire), jusqu'à leur préservation et leur utilisation en tant qu'archives »¹¹⁵. Ce modèle « vise moins les documents eux-mêmes que les manières dont ils sont gérés et leur utilisation potentielle »¹¹⁶. Il s'agit donc d'un modèle théorique de gestion des archives qui considère les documents comme des entités logiques, dynamiques, intégrées dans des contextes spatio-temporels multiples et évolutifs, plutôt que comme des objets statiques et physiques à conserver. « Ces entités sont utilisées dans le temps, d'une manière non linéaire, à des fins transactionnelles de preuve et de mémoire historique »¹¹⁷. Cette approche répond davantage aux besoins du numérique, en permettant un accès continu aux outils

¹¹⁰ RAJOTTE David, « La réflexion archivistique à l'ère du document numérique : un bilan historique », dans *Archives*, vol.42, n°2, 2010-2011, p.82.

¹¹¹ L'archivistique postmoderne, début 21^e siècle, et les courants critiques qui ont suivi ont remis en question et déconstruit les idées de neutralité et d'objectivité traditionnellement associées aux archives, aux archivistes et aux pratiques archivistiques.

¹¹² UPWARD Franck, « Structuring the records continuum, part one : postcustodial principles and properties », dans *Archives and Manuscripts*, vol.24, n°2, 1996, p. 268-285.

¹¹³ LEMAY Yvon et KLEIN Anne, *op.cit.*, p.84.

¹¹⁴ FRINGS-HESSAMI Viviane, « La perspective du Continuum des archives illustré d'un document personnel », dans *Revue électronique suisse de science de l'information*, n°149, 2018, p.3.

¹¹⁵ AS3490-1996, cité par FLYNN Sarah J. A., « The records continuum model in context and its implications for archival practice », dans *Journal of the Society of Archivists*, vol.22, n°1, p.80. Citation traduite avec Deepl.

¹¹⁶ RAJOTTE David, *op.cit.*, p.83.

¹¹⁷ SAAD Nora, ELMQADDEM Nouredine, VASSEUR Edouard, « Archivage à l'ère du numérique : objet et cycle de vie », *op.cit.*, p.183.

d'enregistrement et en assurant la traçabilité des changements des documents au sein de ce même système¹¹⁸.

Le *records continuum* se divise en quatre axes relevant des « préoccupations des archivistes »¹¹⁹ : l'identité (*identity*, qui fait référence « à ceux qui créent ou maintiennent les documents »¹²⁰), la preuve (*evidentiality*, « liées aux documents en tant que preuves intègres et continues »¹²¹), la transaction (*transactionality*, qui « lie le document aux activités »¹²²), et le contenant du document (*recordkeeping container*, « liés aux objets que nous créons afin de stocker les documents »¹²³). Ces axes sont reliés par quatre cercles, chacun représentant une dimension du continuum, relevant des « interventions possibles à réaliser sur un document »¹²⁴ : la création (*create*), la captation (*capture*), l'organisation (*organise*) et la pluralisation (*pluralise*). Les *records* et les archives définitives n'y sont pas considérés comme une étape finale, puisqu'ils peuvent rester soumis à des transformations, portées par de nouveaux usages et de nouveaux contextes, les ramenant de ce fait à la première dimension.

À l'intersection des axes et des dimensions, Upward identifie des concepts « liés à l'utilisation d'un document »¹²⁵.

¹¹⁸ KERN Gilliane, HOLGADO Sandra, COTTIN Michel, *op.cit.*, p.60.

¹¹⁹ RAJOTTE David, *op.cit.*, p.83.

¹²⁰ *Ibid.*

¹²¹ MCKEMMISH Sue e.a. (éd.), *Archives : Recordkeeping in Society*, *op.cit.*, p.202. Citation traduite avec Deepl.

¹²² RAJOTTE David, *op.cit.*, p.83.

¹²³ MCKEMMISH Sue e.a. (éd.), *Archives : Recordkeeping in Society*, *op.cit.*, p.202. Citation traduite avec Deepl.

¹²⁴ RAJOTTE David, *op.cit.*, p.83.

¹²⁵ *Ibid.*

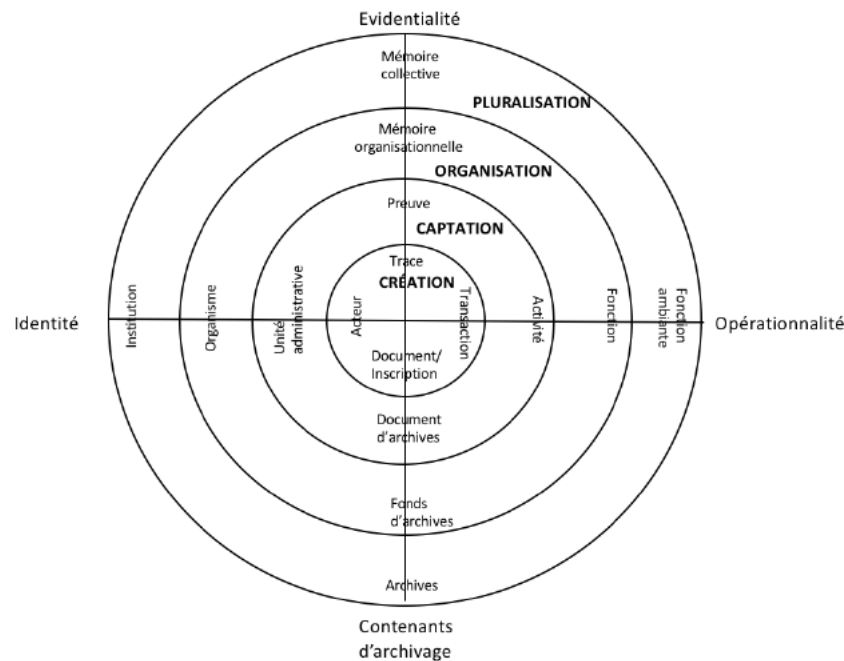


Figure 1 Modèle du records continuum¹²⁶

Les dimensions ne sont pas linéaires, mais simultanées et interreliées. Elles « coexistent parce que les archives sont impactées par les actions de différents acteurs et parce qu'elles peuvent être perçues de manières différentes par des utilisateurs différents »¹²⁷.

Upward explique :

« La première dimension englobe les acteurs qui accomplissent l'acte (décisions, communications, actes), les actes eux-mêmes, les documents qui consignent les actes et la trace, la représentation des actes.

[...] La deuxième dimension englobe les systèmes d'archivage personnels et d'entreprise qui capturent les archives (documents dans leur contexte d'action) de manière à soutenir leur capacité à servir de preuve des activités sociales et commerciales des unités responsables des activités.

[...] La troisième dimension englobe l'organisation des processus d'archivage. Elle concerne la manière dont une personne morale, une organisation ou un individu définit son régime de gestion des documents et, ce faisant, constitue/forme les archives comme mémoire de ses fonctions commerciales ou sociales.

[...] La quatrième dimension concerne la manière dont les archives sont intégrées dans un cadre global (environnemental) afin de fournir une mémoire sociale, historique et

¹²⁶ UPWARD Franck, « Structuring the records continuum, part one : postcustodial principles and properties », *op.cit.*, p. 278, schéma adapté par FRINGS-HESSAMI Viviane, *op.cit.*, p.11.

¹²⁷ FRINGS-HESSAMI Viviane, *op.cit.*, p.3.

culturelle collective des objectifs et rôles sociaux institutionnalisés des individus et des personnes morales »¹²⁸.

En d'autres mots, la dimension *create* désigne le moment où le document est produit dans un contexte donné, avec une intention précise et des métadonnées qui en fixent l'origine. Les concepts qui y sont liés sont la trace, la transaction (ou l'« interaction d'affaire »¹²⁹), les acteurs, le *document* (au sens anglais du terme, donc document non validé).

La dimension *capture* correspond à l'intégration du document dans un système ou un processus qui en garantit la fiabilité, l'authenticité et la traçabilité. Les concepts qui y sont liés sont la preuve, l'activité, l'unité administrative, le *record*.

La dimension *organise* consiste à structurer, classer et relier les documents pour en faciliter la gestion, la compréhension et l'usage dans le temps. Les concepts qui y sont liés sont la mémoire individuelle/organisationnelle, la fonction, l'organisme, le fonds.

La dimension *pluralise* reflète la réutilisation et la circulation des documents dans les contextes multiples, où ils acquièrent de nouvelles significations sociales, culturelles ou historiques. Les concepts qui y sont liés sont la mémoire collective, la fonction d'ambiance (l'utilité), l'institution, les services d'archives.

Ces mêmes concepts peuvent donc être lus aussi selon les axes : l'axe *identity* regroupe les concepts d'acteurs, d'unités, d'organisme et d'institution ; l'axe *transactionality* englobe la transaction, l'activité, la fonction et l'utilité ; l'axe *evidentiality* inclut la trace, la preuve, la mémoire individuelle/organisationnelle, la mémoire collective ; et l'axe *recordkeeping* containers comprend le *document* (au sens anglais du terme), le *record*, les fonds, les services d'archives. Upward résume ainsi la lecture axiale par : « qui (identité) a fait quoi (opérationnalité), quelle preuve existe-t-il à ce sujet (caractère probant) et comment cela peut-il être rappelé à partir de documents d'archives (contenants d'archivage) »¹³⁰.

Bien que le modèle du *records continuum* ait été développé afin de mieux répondre à l'environnement numérique, il est soumis à la critique de certains auteurs. Terry Cook¹³¹, par exemple, estime qu'il n'intègre pas les archives définitives, et en particulier leur exploitation¹³²,

¹²⁸ MCKEMMISH Sue e.a. (éd.), *Archives : Recordkeeping in Society*, op.cit., p.204. Citation traduite avec DeepL.

¹²⁹ LEMAY Yvon et KLEIN Anne, op.cit., p.87.

¹³⁰ UPWARD Frank, « The Records continuum », dans MCKEMMISH Sue e.a. (éd.), *Archives : Recordkeeping in Society*, op.cit., p.202, cité et traduit par KLEIN Anne, *Archive(s), mémoire, art. Éléments pour une archivistique critique*, Laval, Presses universitaires de Laval, 2019, p.107.

¹³¹ COOK Terry, « Beyond the Screen: The Records Continuum and Archival Cultural Heritage », présenté à la Australian Society of Archivists Conference, Melbourne, 18 août 2000, cité par LEMAY Yvon et KLEIN Anne, op.cit., p.90.

¹³² LEMAY Yvon et KLEIN Anne, op.cit., p.89.

contrairement à ce qu'affirment ces promoteurs, à l'instar de Sue McKemmish qui affirme que « modèle du continuum se base sur un concept unificateur des *records*, y compris les archives, qui sont définies comme des *records* ayant une valeur permanente »¹³³. De même, le modèle ne prétend pas rendre compte de l'ensemble des usages sociaux des archives, notamment du point de vue des publics. Luciana Duranti et Heather McNeil formulent également des réserves vis-à-vis du *continuum*. Elles estiment que celui-ci repose sur l'idée d'une continuité homogène dans la gestion des archives, alors que les exigences diffèrent profondément selon les phases classiques des trois âges (voire même des deux âges dans ce cas-ci) : la fiabilité, prioritaire pour les archives courantes, suppose un contrôle rigoureux du processus de création, tandis que l'authenticité, essentielle pour les archives définitives, requiert des méthodes spécifiques de préservation à long terme. Ces objectifs impliquent des approches intellectuelles et techniques distinctes, que le continuum ne prend pas suffisamment en compte selon elles¹³⁴.

5. Conclusion

L'essor du numérique a profondément transformé la pratique archivistique. En effet, les transformations documentaires, techniques et organisationnelles qu'il entraîne conduisent à repenser les notions mêmes d'archives, d'intégrité, de continuité documentaire ou encore de responsabilité professionnelle. Les archivistes ne travaillent plus uniquement avec des objets stables et localisés, mais avec des documents évolutifs, distribués entre plusieurs systèmes, susceptibles d'être modifiés, dupliqués ou réutilisés à différents moments de leur existence.

Dans ce contexte, les modèles archivistiques fondés sur une progression linéaire des archives apparaissent plus difficilement applicables. Le cycle de vie traditionnel, hérité de Pérotin, se heurte à des environnements numériques dans lesquels les temporalités documentaires tendent à se superposer. Les approches du *records management* puis du *records continuum* cherchent dès lors à répondre à cette complexité en intégrant la gestion documentaire dès la création des documents et en envisageant les archives dans une logique plus continue.

Ces évolutions montrent que nombre des questionnements aujourd'hui associés à l'intelligence artificielle s'inscrivent dans des transformations déjà amorcées par le numérique. Toutefois, là où le numérique modifiait principalement les supports, les flux et les dispositifs

¹³³ MCKEMMISH Sue, « Placing records continuum theory and practice », dans *Archival Science*, vol.1, n°4, 2001, reprise par SAAD Nora, ELMQADDEM Nouredine, VASSEUR Edouard, « Archivage à l'ère du numérique : objet et cycle de vie », *op.cit.*, p.182.

¹³⁴ DURANTI Luciana et MCNEIL Heather, « The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project », dans *Archivaria*, n°42, 2001, p.60-61.

de gestion documentaire, l'IA semble intervenir plus directement dans les opérations elles-mêmes, qu'il s'agisse de classer, décrire, extraire, interpréter ou produire de l'information à partir des archives. Dès lors, la question n'est peut-être pas tant de savoir si l'intelligence artificielle constitue une rupture totale avec les évolutions précédentes, mais plutôt de comprendre ce qu'elle ajoute aux reconfigurations déjà engagées par le numérique et quelles nouvelles tensions elle fait émerger autour de l'automatisation, de l'interprétation et de la responsabilité professionnelle.

Deuxième partie. La pratique archivistique face aux enjeux de l'intelligence artificielle

« Nous sommes passés d'une archivistique de fonds à une archivistique de flux. Nous pouvons même parler d'une archivistique computationnelle dans la mesure où les archivistes « sont attirés vers le raisonnement logique des ordinateurs, et les informaticiens et les ingénieurs en logiciel vers les méthodes et principes de l'archivistique » (Jacob et al., 2022) »¹³⁵.

En passant d'une distinction nette et linéaire entre archives courantes, intermédiaires et définitives à une gestion continue des « flux d'information »¹³⁶, le numérique a profondément bouleversé la pratique archivistique. La croissance exponentielle des volumes produits, la dissémination des formats, le brouillage des frontières entre les trois âges et la montée du *records management* ont contribué à déplacer l'intervention archivistique en amont, dans une logique de gouvernance informationnelle et de gestion continue plutôt que séquentielle. Cette reconfiguration a conduit à envisager les archives comme des entités dynamiques, évolutives, en constante transformation plutôt que comme des objets figés dans des étapes successives.

C'est dans ce paysage déjà profondément transformé que s'inscrit aujourd'hui l'intelligence artificielle. Loin de constituer une rupture de même nature que celle induite par le numérique, qui était un changement de support et de temporalité, l'IA apparaît d'abord comme un prolongement des pratiques numériques existantes, telles que l'automatisation des tâches (tri, indexation, capture, vérification), la capacité à absorber le vrac et la masse documentaire, l'amélioration des processus de description ou d'extraction d'informations. Les témoignages abondent dans la littérature scientifique et professionnelle dans le sens d'une modification importante des pratiques archivistiques du fait de l'arrivée de l'IA¹³⁷. Il faut cependant déterminer si ce prolongement constitue un simple approfondissement des dynamiques déjà engagées ou s'il entraîne, à son tour, une reconfiguration plus profonde des modèles archivistiques. L'IA modifie-t-elle les pratiques au point de remettre en cause le cycle de vie traditionnel des archives ou d'exiger un réajustement théorique ? Réoriente-t-elle la lecture du *records continuum* ? Provoque-t-elle un bouleversement comparable, dans la pensée

¹³⁵ SAAD Nora, Elmqaddem NOUREDDINE et VASSEUR Edouard, « L'intelligence artificielle au service de l'archivage numérique : optimisation des pratiques et défis éthiques », dans *Journal of Information Sciences*, vol.3, n°1, 2024, p.69.

¹³⁶ JACOB Steve, SOUISSI Seima et MARTINEAU Charlie, *Intelligence artificielle et transformation des métiers de la gestion documentaire*, Chaire de recherche sur l'administration publique à l'ère numérique, Université de Laval, 2022, p.2.

¹³⁷ *Ibid.*

et la pratique archivistiques, à celui du passage au numérique ? Notons que le propos n'est pas de comparer deux technologies ni deux degrés d'innovation. Le numérique relève en effet principalement d'un changement d'environnement et, de ce fait, des supports, tandis que l'IA introduit davantage un changement dans les modes de traitement, d'analyse, de décision et de production de contenu. Les deux phénomènes ne sont donc pas directement comparables. Il s'agit plutôt ici de comparer leurs effets structurants sur la pratique et la pensée archivistiques, c'est-à-dire la manière dont elles reconfigurent en profondeur les catégories, les méthodes et les repères de la discipline.

Dans la littérature analysée, même anglophone, le cadre d'analyse demeure largement celui du cycle de vie linéaire, à la fois parce qu'il continue d'organiser en pratique la gestion archivistique, comme en témoignent le site internet des Archives générales du Royaume¹³⁸ ou Code du patrimoine français¹³⁹, et parce qu'il structure encore les obligations administratives et les responsabilités des producteurs¹⁴⁰. De plus, une part non négligeable des archives demeure encore aujourd'hui sur support papier, notamment en raison des délais de conservation et de versement, ce qui contribue également au maintien de cette logique séquentielle. En suivant cette logique, qui structure quotidiennement la pratique, nous adoptons nous-mêmes ce schéma classique pour analyser l'impact de l'IA, tout en le mettant en tension avec le *records continuum*, qui permet de penser les archives non en étapes successives mais en dimensions simultanées.

Il convient de rappeler que les principes fondamentaux de l'archivistique (provenance, respect des fonds, intégrité, authenticité) restent non seulement pertinents, mais d'autant plus décisifs dans un contexte où « l'information est de plus en plus éclatée et mouvante »¹⁴¹. En effet, « les deux enjeux considérés comme les plus importants s'ancrent parfaitement dans notre ère digitale : 88 % des archivistes placent la préservation de l'authenticité et de l'intégrité des archives numériques au cœur de leurs préoccupations, tandis que la gestion de la masse croissante de données mobilise 76 % des professionnels. La sécurisation des données sensibles

¹³⁸ « Conserver de l'information », sur *Archives Générales du Royaume*, <https://www.arch.be/index.php?l=fr&m=fonctionnaire&r=conserver-de-l-information>, consulté le 24 juin 2025.

¹³⁹ R.212-10 à R.212-12 du Code du Patrimoine français.

¹⁴⁰ GUYON Céline, « Théorie, technique et pratique archivistique en environnement numérique. Colloque International sur le Document Electronique : Document et archivage : pratiques formelles et informelles dans les organisations », *CIDE23*, 2023, Grenoble, France, p. 5, sur <https://hal.science/hal-04371177v1>, consulté le 24 juin 2025.

¹⁴¹ COUTURE Carol, « La discipline archivistique au Canada : état de développement et perspectives d'avenir », sur *In Situ*, n°30, 2016, p.2, <http://journals.openedition.org/insitu/13669>, consulté le 24 juin 2025.

(67 %) et l'adaptation aux nouvelles technologies (63 %) représentent aussi des thèmes majeurs pour le futur des archivistes, sans oublier l'évolution du cadre juridique et réglementaire »¹⁴².

C'est à l'examen de ces transformations, de leurs implications théoriques et de leurs effets sur les pratiques archivistiques que ce chapitre sera consacré. Pour appréhender ces transformations, le chapitre adopte une progression similaire à celle de la partie consacrée au numérique. Nous examinerons d'abord ce que l'IA engendre sur la forme documentaire, en mentionnant quelques projets archivistiques concrets. Nous analyserons ensuite ses effets sur les dispositifs de gestion et de conservation, avant d'observer la manière dont elle reconfigure les gestes professionnels au quotidien. La mise sous tension du cycle de vie des archives terminera enfin ce chapitre.

Selon une étude menée par *Archimag*, publiée en février 2026, « 13 % des archivistes [interrogés] se déclarent comme étant des utilisateurs réguliers d'IA et 40 % des utilisateurs ponctuels. Un tiers des professionnels souhaiterait l'exploiter davantage dans son métier, contre 14 % qui n'en voient pas l'intérêt »¹⁴³. Il apparaît que pour 73% de la cohorte étudiée « les nouvelles technologies (de manière générale) vont engendrer des changements importants, mais graduels, d'ici cinq ans. Selon 15 % de notre panel, elles vont impacter très profondément et totalement le métier, contre 13 % qui n'envisagent peu ou pas de transformations »¹⁴⁴. Le magazine a recensé les différents usages de l'IA par les archivistes prenant part à l'étude (voir figure ci-dessous). 64% des archivistes interrogés font usage de l'IA générative, dans le cadre de la production de résumés et de synthèses, de recherche d'information ou de rédaction de contenus¹⁴⁵, les différentes potentialités de l'IA restent donc assez peu connues ou mobilisées.

¹⁴² SIVAGAMI Casimir, « Archiviste : enquête sur un métier en constante évolution », dans *Archimag*, n°391, février 2026 p.23.

¹⁴³ *Ibid.*

¹⁴⁴ *Ibid.*

¹⁴⁵ JOST Clémence, « Le baromètre des usages des IA génératives par les métiers de l'information », dans *Archimag. Guide pratique n°83. IA génératives : cas d'usage et retours d'expérience*, février 2025, p.5.

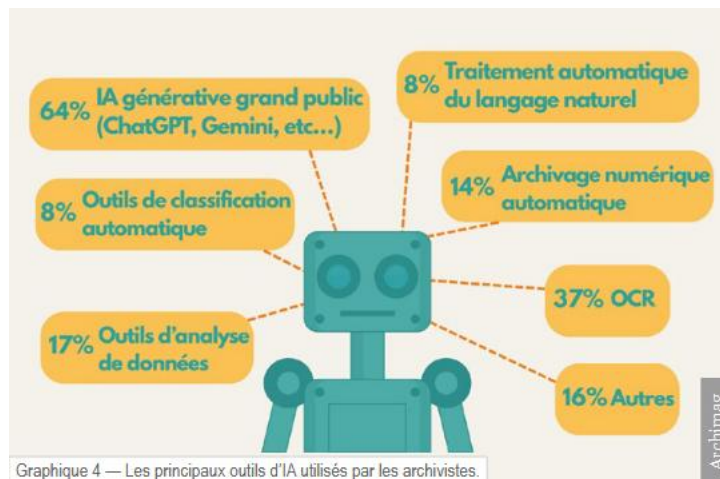


Figure 2 Les principaux outils d'IA utilisés par les archivistes¹⁴⁶

L'intelligence artificielle ne vient pas bouleverser de fond en comble les pratiques archivistiques, puisqu'elle prolonge surtout un mouvement déjà engagé avec le numérique. La dématérialisation avait fait apparaître des documents plus instables, des opérations imbriquées, et la nécessité pour les archivistes de gérer les traces, les versions et les contextes de manière continue¹⁴⁷. L'IA accentue cette logique en rendant certaines opérations plus rapides, plus massives et parfois simultanées. Elle accélère une dynamique déjà présente, celle d'archives en flux, qui se créent, se transforment et se réutilisent en continu.

1. L'IA et la forme documentaire

Les transformations induites par l'IA dans les organisations touchent en premier lieu le document lui-même, car elle ajoute des couches analytiques pour produire ses couches descriptives, son accessibilité, ses métadonnées et les éléments qui permettent de l'interpréter et de l'inscrire dans un contexte archivistique. Alors que le numérique avait déjà modifié la fixité des documents, leur stabilité formelle et le lien archivistique, l'IA ajoute donc un niveau supplémentaire de complexité : elle crée des transcriptions, extrait des entités, génère des mots-clés, produit des résumés, recompose la structure logique d'un document, et peut même en déduire des catégories ou des relations. Ces opérations automatisées augmentent la forme documentaire, la déplacent, et parfois la brouillent, en produisant des représentations computationnelles, c'est-à-dire opérables par l'IA, qui coexistent avec l'objet initial, sur base des cas d'analyse concrets repérés dans la littérature professionnelle. Nous avons repéré et postulons que cette dynamique est perceptible dans trois grands types d'opérations, à savoir la

¹⁴⁶ *Ibid.*

¹⁴⁷ JACOB Steve, SOUISSI Seima et MARTINEAU Charlie, *op.cit.*, p.6.

mise en lisibilité de documents complexes pour l'IA (notamment manuscrits, qui doivent être décryptés, transcrits et structurés), l'extraction d'information et la génération de métadonnées, puis la structuration computationnelle du contenu. Nous verrons ainsi quelques exemples de comment l'IA rend lisible des documents complexes, en enrichit la description et produit des formes documentaires augmentées.

a. Un accès renouvelé aux contenus textuels et manuscrits

L'IA permet aujourd'hui d'accéder à des corpus textuels variés, tels que des manuscrits, documents hétérogènes, des fonds volumineux, ou encore des dossiers structurés de manière irrégulière, qui étaient auparavant difficilement exploitables. À ce titre, l'exemple du projet européen *Himanis* est particulièrement significatif. Celui-ci mobilise des modèles d'apprentissage automatique pour indexer plus de 80 000 pages de registres médiévaux, caractérisés par des mises en page irrégulières, en latin et en français et une grande variabilité graphique. Sans automatisation, de tels corpus seraient difficilement exploitables, en raison de la technicité paléographique qu'ils requièrent et de leur volume. Grâce à l'IA, ils deviennent interrogeables puisqu'il devient possible de rechercher des mots, de naviguer dans les contenus ou d'identifier des informations récurrentes¹⁴⁸. Un autre exemple se trouve dans le travail autour de *Transkribus*, utilisé notamment pour le fonds Louis-François-Georges Baby, qui met en évidence la capacité de l'IA à produire des transcriptions pouvant être relues, corrigées ou enrichies par les archivistes.¹⁴⁹ Dans la même veine, le projet *Socface*, mené par le Service interministériel des archives de France, traite à très grande échelle des listes nominatives de recensement de la population entre 1836 et 1926. Après numérisation, les documents sont « transcrits automatiquement et analysés pour constituer une base de données de plusieurs millions d'individus »¹⁵⁰.

Ces transcriptions ouvrent également la voie à une révision critique des descriptions archivistiques existantes, notamment lorsqu'il s'agit d'identifier, par exemple des termes sensibles, des contextes coloniaux ou des catégories historiquement marquées¹⁵¹. Ces technologies redéfinissent les conditions d'accès aux différents contenus. Elles leur ajoutent

¹⁴⁸ CASTANIÉ Camille, « L'intelligence artificielle permet de nouveaux projets pour les archivistes », dans *Archivistes !*, n°154, octobre-décembre 2025, p.21.

¹⁴⁹ TEXIER Bruno, « IA et archives : du patrimoine culturel aux documents d'entreprise », dans *Archimag. Guide pratique* n°77. *L'IA générative : l'utiliser dans ses projets professionnels*, juin 2024, p.26-27.

¹⁵⁰ *Ibid.*

¹⁵¹ ID., « Université de Montréal : l'IA pour déchiffrer des documents manuscrits », dans *Archimag*, n°380, décembre 2024-janvier 2025, p.16.

BERNY Camille, *Essai : Les biais descriptifs dans les archives. Vers une pratique éthique et inclusive. Travail rédigé dans le cadre du cours LARGD2301 – Enjeux sociétaux, éthiques et questions d'actualité*, UCLouvain, août 2025.

une transcription ou une version textuelle qui en facilite la lecture, la recherche et l'utilisation, tout en élargissant potentiellement le public pouvant y accéder. Une transcription HTR devient une nouvelle « couche », un objet documentaire dérivé susceptible d'être indexé, annoté, interconnecté à d'autres documents via des entités ou des identifiants. En somme, l'IA permet de rendre certains corpus historiques plus lisibles, mais cette « lisibilité computationnelle », opérable, reste un produit dérivé puisqu'elle ouvre l'accès aux contenus, tout en introduisant une médiation algorithmique de l'IA qui doit être expliquée, contrôlée et replacée dans son contexte par l'archiviste¹⁵².

b. Extraction d'information et enrichissement automatique des métadonnées

Au-delà de la transcription, l'une des contributions majeures de l'IA consiste à extraire des informations pertinentes et à enrichir automatiquement les métadonnées associées aux documents¹⁵³. Plusieurs outils combinent reconnaissance de contenu, extraction de métadonnées natives, détection d'entités nommées (NER), classification documentaire et analyse sémantique¹⁵⁴. Bien que dans la pratique, « les résultats restent mitigés, surtout face aux données non structurées qui dominent les fonds »¹⁵⁵, ces fonctionnalités permettent d'identifier des personnes, des lieux ou des dates dans des contenus textuels, d'attribuer automatiquement une typologie documentaire (contrat, facture, rapport, etc.), de segmenter des textes complexes en unités cohérentes, de proposer des mots-clés ou des catégories pertinentes. L'extraction automatique peut également jouer un rôle important dans la gestion du vrac documentaire, notamment dans les messageries électroniques ou les espaces collaboratifs. À titre d'exemple, le programme *Pêle-mél*, mené par l'Université d'Angers, l'Ecole des Chartes et le ministère de la Santé français, utilise des techniques de traitement automatique du langage (TALN) pour traiter et classifier des volumes massifs d'e-mails à archiver, « d'aider l'archiviste à interpréter et à contextualiser une ou des boîtes aux lettres via l'extraction d'entités nommées, de termes »¹⁵⁶. Ce type d'outil transforme la forme documentaire en la rendant plus structurée,

¹⁵² Pour cette réflexion, l'outil *Copilot* a été utilisé en tant qu'aide à « l'exploration d'idées », sur la façon dont l'IA modifie et enrichit la forme documentaire d'un document retranscrit, parallèlement au développement consacré à l'effet du numérique sur la forme documentaire. La réflexion a été alimentée par les articles de JACOB Steve, SOUSSI Seima et MARTINEAU Charlie, *op.cit.* et ALAOUI Siham, *op.cit.*

¹⁵³ VÉROVE Simon et BAHIER MICHEL Livia, « Réflexion sur les apports de l'IA au métier. IA : vers une réalité archivistique ? », dans *Archivistes !*, n°47, janvier-mars 2024, p.33-34.

¹⁵⁴ REMIZE Michel, « Automatisation : aller encore plus loin avec l'IA générative », dans *Archimag*, n°380, décembre 2024-janvier 2025, p.26-27.

¹⁵⁵ DEBEAUVAIS Jeanne, « Intégrer l'IA dans la gestion documentaires : un retour d'expérience », dans *Archivistes !*, n°155, janvier-mars 2026, p.26.

¹⁵⁶ « Journée d'étude AAF – Ecole des Chartes. L'intelligence artificielle dans la gestion des données », sur AAF. Association des Archivistes Français, <https://www.archivistes.org/ressources/journee-detude-aaf-ecole-des-chartes-lintelligence-artificielle-dans-la-gestion-des-donnees/> (consulté le 12 décembre 2025).

plus lisible, en produisant pour ce faire des métadonnées en grande quantité¹⁵⁷. En outre, grâce à l'extraction de contenu documentaire, des outils permettent de proposer des durées de conservation pour les documents, bien que « l'archiviste reste maître des choix en termes d'organisation des données et d'évaluation archivistique déterminant la conservation ou la destruction »¹⁵⁸.

Ce mouvement d'enrichissement modifie la forme documentaire dans au moins trois dimensions. D'une part sur le niveau d'échelle, puisque l'unité documentaire peut être réduite au *token*, dans le but d'un traitement par l'IA, d'autre part sur la contextualisation, où les entités nommées ajoutent des liens logiques entre des ensembles documentaires, et enfin sur la densité descriptive, puisque le travail par l'IA ajoute des métadonnées supplémentaires¹⁵⁹.

c. Une structuration computationnelle du document : vers une forme documentaire augmentée

Comme dit plus haut, l'IA produit des représentations dérivées qui transforment la manière dont les documents sont repérés ou mobilisés dans les systèmes. Ces représentations computationnelles, sous forme *d'embeddings*, de résumés automatiques, d'entités nommées, etc., ne constituent pas des documents en tant que tels, mais des couches supplémentaires qui viennent s'ajouter aux contenus d'origine et participent à une forme documentaire augmentée. Ainsi, elles permettent de naviguer dans des corpus volumineux, d'identifier des similarités conceptuelles ou de faciliter la recherche thématique, tout en introduisant des interprétations produites statistiquement.

Dans ce contexte, les *paradata* jouent un rôle essentiel. Celles-ci sont définies comme des « *information about the procedure(s) and tools used to create and process information resources, along with information about the persons carrying out those procedures* »¹⁶⁰, elles constituent une véritable documentation processuelle. Elles sont indispensables pour garantir la transparence, l'explicabilité et la responsabilité des traitements algorithmiques, mais également

¹⁵⁷ SIVAGAMI Casimir, « Pêle-mêle pour l'archivage des e-mails », dans *Archimag. Guide pratique n°83. IA générative : cas d'usage et retours d'expérience*, p.40.

¹⁵⁸ Echange par email avec Johan Pierret, délégué à l'archivage au Ministère de la Santé et de la Sécurité sociale du Grand-Duché du Luxembourg, 25 juillet 2025.

¹⁵⁹ Pour cette réflexion, l'outil *Copilot* a été utilisé en tant qu'aide à « l'exploration d'idées », sur la façon dont l'IA modifie et enrichit la forme documentaire d'un document retranscrit, à partir des lectures de JACOB Steve, SOUSSI Seima et MARTINEAU Charlie, *op.cit.* et BASMA Makhoul Shabou, « Humanités numériques & Sciences de l'information : ressemblance, complémentarité et développements mutualisés », dans *Journal of information sciences*, vol.1, n°21, 2022, p.82-83.

¹⁶⁰ DAVET Jeremy, HAMIDZADEH Babak, FRANKS Patricia, « Archivist in the machine: paradata for AI-based automation in the archives », dans *Archival Science*, vol.23, 2023, p.276.

pour maintenir l'authenticité, l'intégrité et la fiabilité des documents enrichis¹⁶¹. Les *paradata* apparaissent ainsi complémentaires des métadonnées. En effet, tandis que les métadonnées renseignent le document et son contexte, les *paradata* documentent davantage les processus, outils et interventions ayant conduit à sa création, son traitement ou son enrichissement.

d. Conditions, biais et limites

L'intégration de l'IA dans la pratique archivistique s'accompagne d'un ensemble de limites qui encadrent et relativisent ses apports. Les contraintes ne relèvent pas seulement de difficultés techniques (liées à la machine ou au corpus), mais touchent aussi à des dimensions éthiques, juridiques et méthodologiques, qui posent la question de la qualité et de la fiabilité des transformations opérées sur les documents. Un premier ensemble de limites concerne la fiabilité des productions des modèles. L'IA peut générer des hallucinations, c'est-à-dire produire des réponses plausibles, mais incorrectes, ou effectuer des erreurs d'interprétation lorsque les données d'entrée ne correspondent pas à ce pour quoi elle a été entraînée. Les modèles peuvent en effet reproduire les lacunes de leurs jeux d'entraînement¹⁶². Rappelons que l'IA ne comprend pas les documents qu'elle traite, elle les interprète à partir d'opérations statistiques, de mécanismes de comparaison et, dans certains cas, du recours à des outils ou modules de calcul externes. Ces biais sont renforcés par un facteur technique déterminant, à savoir la qualité des données en entrée, qui dépend de la structure des documents, de leur homogénéité, voire de leur numérisation. Face à ces limites, certains garde-fous peuvent être mis en place par les professionnels. L'archiviste demeure responsable de la qualité, de la lisibilité et de la contextualisation des documents, ainsi que de la mise en œuvre et de la supervision des traitements, tandis que l'IA intervient principalement comme outil d'assistance et d'automatisation¹⁶³.

À cela s'ajoutent les implications de protection des données personnelles, puisque l'extraction automatique de données personnelles des archives induit immédiatement les exigences légales de minimisation des données, de protection, de documentation des opérations et de contrôle de l'accès¹⁶⁴.

¹⁶¹ DURANTI Luciana et ROGERS Corinne, « Artificial Intelligence and Documentary Heritage », Memory of the World Sub on Education and Research (SCEaR), Newsletter 2024. Special Issue 2024, p.58 et BASMA Makhoulf Shabou, « Fonction d'évaluation des archives. Bilan sommaire des développements, des enjeux actuels et des défis futurs », dans GAGNON-ARGUIN Louise et LAJEUNESSE Marcel (éd.), *op.cit.*, p.200.

¹⁶² JOST Clémence, « Nicolas Curien : le rapport à la vérité est mal parti avec les IA génératives », dans *Archimag*, n°385, juin 2025, p.48-49.

¹⁶³ ALAOUI Siham, *op.cit.*, p.11-17.

¹⁶⁴ GRINBAUM Alexei e.a., *op. cit.*, p. 27-30.

e. Limites théoriques : métadonnées, authenticité et statut du record à l'ère de l'automatisation

L'IA, en produisant des représentations dérivées et en automatisant une partie de la description, introduit certaines limites qui tiennent à la provenance des métadonnées générées, à la possibilité de démontrer l'authenticité du document lorsque les traitements sont partiellement opaques, et à la manière dont la fixité du document peut être maintenue dans un environnement où les transformations computationnelles se multiplient. Ces limites renvoient en partie à l'absence de documentation systématique des traitements, d'où la nécessité de documenter systématiquement les traitements, dans une logique de *paradata*, afin de garantir la lisibilité des opérations et de répondre aux exigences d'*accountability*, largement soulignées dans les travaux récents sur l'usage de l'IA en contexte archivistique¹⁶⁵. Le principe d'*accountability*, d'IA responsable, est définie comme « *the ability to explain and provide evidence of the actions or decisions made by a system of interlocking elements that accept input, process data using an algorithm and model, and produce a result. It is also the ability to take responsibility for the outcomes and impacts of an AI system. It further involves the ability to monitor, audit, and correct the system if it deviates from its intended purpose or causes harm* »¹⁶⁶.

D'une part, la frontière entre métadonnées archivistiques et métadonnées produites par IA devient moins lisible. Les métadonnées générées (entités nommées, mots-clés, classifications automatiques) peuvent être identiques dans leur forme à celles élaborées par les archivistes, mais elles n'ont pas la même provenance, puisqu'elles résultent d'un calcul probabiliste dépendant d'un modèle, et non d'une intention descriptive de l'archiviste. Faute de documenter cette provenance computationnelle au moyen des *paradata*, elles risquent d'être intégrées comme des métadonnées professionnelles, brouillant le lien archivistique et introduisant des interprétations qui ne relèvent pas du contexte de création du document¹⁶⁷.

D'autre part, les opérations algorithmiques produisent des représentations dérivées qui facilitent l'accès aux documents, mais introduisent une médiation invisible. Sans

¹⁶⁵ FRANKS Patricia, « In the pursuit of archival accountability : positioning paradata as AI processual documentation », *Society of American Archivists, Research Forum 2023*, https://www2.archivists.org/sites/all/files/Franks_In%20the%20Pursuit%20of%20Archival%20Accountability.pdf (consulté le 15 janvier 2026).

¹⁶⁶ ID., « In the pursuit of archival accountability : positioning paradata as AI processual documentation », *Society of American Archivists, Research Forum 2023*, https://www2.archivists.org/sites/all/files/Franks_In%20the%20Pursuit%20of%20Archival%20Accountability.pdf (consulté le 15 janvier 2026).

¹⁶⁷ FELICIATI Pierluigi et DURANTI Luciana, « The responsible use of artificial intelligence in archives through the use of paradata », dans *JLIST.it*, vol.6, n°2, mai 2025, p.6.

contextualisation, ces dérivés peuvent être pris pour partie intégrante du document original, fragilisant la démonstration de l'authenticité¹⁶⁸. Les recherches menées dans le cadre d'*InterPARES* Trust AI soulignent que la documentation des processus, grâce aux *paradata*, est nécessaire pour rendre visibles ces transformations, notamment lorsque les traitements sont difficilement explicables ou non reproductibles à l'identique¹⁶⁹.

Enfin, l'IA met sous tension la fixité documentaire. En fragmentant les documents (entités, segments, résumés), en les enrichissant de manière continue et en réinterprétant leur contenu (typologies, catégories, données sensibles), l'IA multiplie les versions et les dérivés susceptibles de coexister dans les systèmes. Ces transformations successives ne suppriment pas la pertinence de la notion de fixité, mais elles exigent une documentation renforcée pour préserver la lisibilité du document, son identité et sa valeur probante¹⁷⁰.

Ces limites théoriques rappellent que l'intégration de l'IA dans la chaîne documentaire doit s'accompagner de garanties en matière de provenance, de traçabilité et de supervision professionnelle, afin de maintenir la cohérence des documents et les principes fondamentaux de l'archivistique.

2. L'IA et les dispositifs de gestion et de conservation

L'intégration de l'IA dans les systèmes documentaires modifie aussi les dispositifs de gestion et de conservation dans lesquels les documents circulent et sont pris en charge. La littérature professionnelle montre que l'IA s'inscrit aujourd'hui dans les processus documentaires, les chaînes de traitement, les outils de capture, les systèmes d'archivage et les cadres réglementaires, devenant un élément structurant de la gouvernance informationnelle¹⁷¹.

a. Automatisation des processus documentaires

Dans de nombreuses organisations, l'IA est mobilisée pour automatiser ou accélérer des opérations auparavant chronophages, tels que le moissonnage, la détection de doublons, l'extraction, ou encore le classement automatique. Ces usages apparaissent particulièrement dans le traitement du vrac numérique, qu'il s'agisse d'espaces SharePoint, de dossiers

¹⁶⁸ *Ibid.*

¹⁶⁹ DAVET Jeremy e.a., « Tracking the Functions of AI as Paradata & Pursuing Archival Accountability », dans *Archiving 2021. Final program and proceedings*. Society for imaging science and technology, p.83-88.

¹⁷⁰ DAVET Jeremy, HAMIDZADEH Babak et FRANKS Patricia, « Archivist in the machine : paradata for AI-based automation in the archives », dans *Archival Sciences*, vol.23, 2023, p.278-281.

¹⁷¹ Cf. ALAOU SIHAM, *op.cit.*

collaboratifs, ou de boîtes mail volumineuses¹⁷². Au Service des Archives de l'UCLouvain, Delphine Picon traite ainsi un versement de plus de 500 GB de données numériques à archiver. L'IA est mobilisée afin d'aider à l'implémentation d'une politique de traitement définie en amont par l'archiviste, notamment par la création de scripts Python permettant d'établir un inventaire de versement, de supprimer certains doublons et de gérer la question de l'obsolescence des extensions¹⁷³. Cette pratique du « no-code » semble particulièrement se développer puisqu'elle permet aux archivistes, avec peu ou pas de bagage en programmation, de « développer de manière autonome des solutions métiers sûres et spécifiques à leurs besoins »¹⁷⁴.

b. IA et conservation : rétention, contrôle, conformité

L'IA intervient ensuite dans la conservation des documents, en particulier sur les questions de rétention, de contrôle qualité et de conformité. Certains outils permettent, par exemple, de suggérer automatiquement des durées de conservation, sur la base d'une analyse du contenu¹⁷⁵ ou de la typologie documentaire, mais aussi du cadre réglementaire applicable. Ces opérations demeurent toutefois particulièrement sensibles dans la mesure où les décisions de destruction présentent un caractère définitif et engagent directement la responsabilité des institutions et des archivistes. Ces suggestions ne se substituent pas au jugement professionnel, mais peuvent constituer une aide au tri et à la planification des destructions ou des versements¹⁷⁶.

L'IA peut également aider au contrôle qualité archivistique, en détectant des anomalies dans les métadonnées, en identifiant des données personnelles ou sensibles, ou en repérant des documents essentiels. Cette intégration dans les dispositifs documentaires s'inscrit dans un ensemble de cadres réglementaires¹⁷⁷. Ainsi, le RGPD demeure le texte de référence dès qu'une extraction, une classification automatique ou une analyse de contenu implique des données personnelles. Le règlement impose les principes de finalité, de minimisation, de proportionnalité, de sécurisation, mais aussi de documentation des opérations (registre des traitements, base légale, justification des durées de conservation). Toute automatisation impliquant des données personnelles doit donc être évaluée en regard de ces principes, y

¹⁷² Echange par email avec Johan Pierret, délégué à l'archivage au Ministère de la Santé et de la Sécurité sociale du Grand-Duché du Luxembourg, 25 juillet 2025.

¹⁷³ Entretien avec Delphine Picon archiviste à l'UCLouvain, le 16 septembre 2025.

¹⁷⁴ Echange par email avec Johan Pierret, délégué à l'archivage au Ministère de la Santé et de la Sécurité sociale du Grand-Duché du Luxembourg, 25 juillet 2025.

¹⁷⁵ SIVAGAMI Casimir, « Stéphane Wulc : la complexité comme terrain de jeu », dans *Archimag*, n°380, décembre 2024-janvier 2025, p.43.

¹⁷⁶ ALAOUI Siham, *op. cit.*, p.7.

¹⁷⁷ ALAOUI Siham, *op. cit.*, p.10.

compris lorsqu'elle intervient en amont, au moment du tri, de la pré-classification ou de la détection de données sensibles¹⁷⁸.

Le règlement européen sur l'intelligence artificielle (*AI Act*), pour sa part, introduit une classification par niveaux de risque en classant les systèmes d'IA et les usages qui en sont faits : risque inacceptable, haut risque, risque limité, risque minimal¹⁷⁹. Certains usages jugés incompatibles avec les droits fondamentaux, comme certaines formes de traitements biométriques, peuvent ainsi être interdits. Des modèles dont les usages sont classés à haut risque sont soumis à des obligations renforcées, telles que celles de transparence, de documentation et de conservation des preuves de fonctionnement, d'évaluations préalables et continues, d'auditabilité et de supervision humaine obligatoire¹⁸⁰. Pour répondre aux exigences de l'*AI Act*, le principe d'explicabilité (XAI) devient d'ailleurs primordial afin de justifier le raisonnement algorithmique auprès des auditeurs¹⁸¹.

Enfin, bien qu'il ne vise pas directement l'intelligence artificielle, le règlement eIDAS 2, qui redéfinit au niveau européen les conditions de l'archivage électronique qualifié, encadre les services de confiance, et particulièrement l'archivage électronique qualifié. Ses exigences portent sur l'intégrité, la preuve, la sécurité et la traçabilité des documents conservés¹⁸². Si une organisation utilise des traitements algorithmiques, que ce soit de la retranscription, de l'extraction, ou encore de la classification, avant versement, elle doit être capable de documenter comment les données ont été transformées, par quels systèmes, avec quelle fiabilité, et selon quelles règles. Ainsi, les *paradata* et l'XAI peuvent être garantes de l'origine et de la validité des opérations précédant le versement.

c. Recherche enrichie et nouvelles formes de valorisation

Comme vu plus haut, les dispositifs d'accès exploitent l'IA, permettant une meilleure valorisation par la recherche et l'exploration des fonds qui sont conservés. C'est d'ailleurs cette facette de l'IA qui est principalement traitée dans la littérature, les colloques et événements mêlant archives et IA. À titre d'exemple, l'INA utilise la transcription et l'analyse automatique pour améliorer la « découvrabilité » de ses archives audiovisuelles, tout en intégrant des mécanismes de contrôle qualité et en tenant compte des contraintes juridiques, notamment en

¹⁷⁸ GRINBAUM, *op. cit.*, p. 23.

¹⁷⁹ GRINBAUM, *op. cit.*, p. 22.

¹⁸⁰ Réf IA Act

¹⁸¹ Saad Nora, p. 85

¹⁸² TEXIER Bruno, « L'archivage électronique au temps d'eIDAS 2 », dans *Archimag*, n°387, septembre 2025, p.24.

matière de droits d'auteur.¹⁸³ La Philharmonie de Paris mobilise également l'IA pour transcrire et indexer en profondeur ses « archives de la parole », en extrayant des entités musicales, en générant des *embeddings* et en proposant des visualisations interactives pour faciliter la recherche¹⁸⁴. La SNCF, enfin, déploie un *chatbot* conçu pour rendre accessibles « 50 kilomètres d'archives historiques et plus de 450 000 photographies conservées par le Service Archives et Documentation »¹⁸⁵.

3. L'IA et les pratiques professionnelles

Enfin, l'intégration de l'IA dans les systèmes documentaires modifie également, et parfois en profondeur, les pratiques professionnelles des archivistes et des gestionnaires de l'information. Cette transformation ne signifie ni un effacement des métiers ni une substitution de l'humain, mais bien un déplacement de leur agentivité, leur capacité d'agir, allant de l'exécution vers le contrôle et le pilotage de l'IA¹⁸⁶.

a. Évolution des gestes : de l'exécution au pilotage

La littérature et les professionnels témoignent de la même idée et de l'importance pour l'archiviste de déplacer son métier de l'exécution vers la supervision. Les opérations répétitives (tri, classification, extraction, indexation) peuvent désormais être partiellement automatisées¹⁸⁷. Chaque étape nécessite cependant une vérification humaine, qui devient un geste professionnel à part entière. Ainsi, l'archiviste demeure indispensable « pour alimenter et vérifier la qualité des algorithmes, valider une décision prise par un algorithme ou résoudre des problèmes ou encore traiter les cas difficiles »¹⁸⁸. Cette supervision prend plusieurs formes. D'abord, l'archiviste vérifie les résultats produits par l'IA. Ainsi, une classification automatique, une extraction d'entités nommées, une proposition de durée de conservation ou un résumé généré doivent être contrôlés pour en évaluer la cohérence. Rappelons-le, les modèles peuvent produire des erreurs d'interprétation, de faux positifs ou des données incomplètes. Ensuite, l'archiviste

¹⁸³ « Comment l'INA met l'IA au service de la valorisation du patrimoine audiovisuel », dans *Archimag*, supplément gratuit au n°367, juillet 2024, p.7.

¹⁸⁴ TEXIER Bruno, « l'IA en mode majeur à la Philharmonie de Paris », dans *Archimag*, n°380, décembre 2024-janvier 2025, p.19.

¹⁸⁵ KAËLIG Alléaume, « SNCF : un chatbot embarque les visiteurs sur les rails du patrimoine », dans *Archimag*, n°380, décembre 2024-janvier 2025, p.20.

¹⁸⁶ Cf ZHANG Alexander et VARSHNEY Lav R., « Conceptualizing Agency : A Framework for Human-AI Interaction », dans *CEUR Workshop Proceedings*, n°3957, 2025, <https://experts.illinois.edu/en/publications/conceptualizing-agency-a-framework-for-human-ai-interaction/> (consulté le 13 janvier 2025).

¹⁸⁷ ALAOUI Siham, *op.cit.*, p. 5.

¹⁸⁸ JACOB Steve, SOUISSI Seima et MARTINEAU Charlie, *op.cit.*, p.8.

intervient comme garant de la cohérence documentaire¹⁸⁹. Dans les processus documentaires automatisés, il doit confirmer que les règles de classement, d'indexation et de conservation ont été correctement appliquées. Cette vigilance est indispensable pour prévenir les erreurs de tri pouvant entraîner une destruction non désirée ou une rétention excessive¹⁹⁰.

Enfin, l'IA introduit un besoin nouveau, à savoir, la correction des biais, qui peuvent générer des erreurs ou discrimination. Les modèles reproduisent les limites des données d'entraînement, et l'archiviste doit en identifier les effets, repérer des catégorisations problématiques ou des descriptions trop partielles, et ajuster les paramètres ou les règles¹⁹¹. Le geste professionnel n'est donc plus seulement celui qui accomplit une tâche, mais celui qui oriente, supervise et valide l'activité des systèmes automatisés. Cette redéfinition renforce l'importance éthique de « l'humain dans la boucle » : l'archiviste agit comme un acteur responsable du sens, de l'intégrité et des conséquences des opérations algorithmiques.

b. Nouvelles compétences et collaborations étendues

La transformation des gestes s'accompagne d'une évolution des compétences, qui prend plusieurs formes, que ce soit par le *prompting* et les compétences en technologie, l'analyse des risques, ou encore l'analyse des *paradata*. Ces nouvelles compétences entraînent également une reconfiguration des collaborations, qui s'accroissent peut-être plus qu'auparavant. Que ce soit avec les informaticiens, les juristes, les *data stewards*, ou encore les *data librarian*. Ces collaborations participent à l'émergence d'une pratique archivistique élargie, orientée vers la gouvernance informationnelle, dans laquelle l'archiviste endosse un rôle transversal¹⁹².

c. L'IA comme intensification du mouvement vers le *records management*

L'IA ne crée pas une rupture conceptuelle dans les pratiques archivistiques, mais elle accentue un mouvement déjà engagé. On peut en effet voir le renforcement du déplacement en amont des interventions archivistiques, vers les moments mêmes où les documents sont

¹⁸⁹ ALAOUI Siham, *op. cit.*, p.12.

¹⁹⁰ Service des ressources informationnelles et archives de l'Université de Lausanne, « Records management et gouvernance informationnelle : Concepts et explications », 2016, p.5-6, https://www.unil.ch/files/live/sites/unil/files/02-universite/0204-organisation/020404-unites-et-services/UNIRIS/references/UNIL_RM_Gouvernance_Info_concepts_explications_VF-1.pdf, consulté le 17 juin 2025.

¹⁹¹ BASMA Makhoul Shabou, « Archivistique à l'ère de l'IA : opportunités, défis et besoin d'utilisation responsable », dans *Arbido, Les archives du futur*, vol. 1, 2023, <https://arbido.ch/fr/edition-article/2023/archives-du-futur/archivistique-a-ler-de-lia-opportunités-défis-et-besoin-dutilisation-responsable>

¹⁹² KAËLIG Alléaume, « Data librarian, au cœur du cycle de vie des données de la recherche », dans *Archimag*, n°387, septembre 2025, p.18 et ALAOUI Siham, *op. cit.*, p.12-17, JOST Clémence, « Nicolas Curien : le rapport à la vérité est mal parti avec les IA génératives », dans *Archimag*, n°385, juin 2025, p.48-49, International Council on Archives, *AI and Archival Practice – Online Tutorials. Session 2: AI and Digital Preservation*, 2024, <https://www.ica.org/resource/ai-and-archival-practice-on-line-tutorials/> (consulté le 19 août 2025).

produits, reçus ou capturés. Dans la pratique quotidienne, l'automatisation engendrée par l'IA donne l'impression que certaines opérations, comme la pré-classification, la détection de données personnelles, ou la vérification de cohérence, se déclenchent plus tôt qu'auparavant. L'intervention de l'archiviste est primordiale dans une « perspective de *preservation by design* »¹⁹³, ou « *archiving by design* »¹⁹⁴. Ce dernier principe se définit comme « une méthode pour assurer l'accessibilité durable des documents dès le moment de la création ou de la capture. Les principes fondamentaux de l'accessibilité durable incluent Trouvable, Disponible, Lisible, Interprétable, Fiable et Durable/à l'épreuve du temps »¹⁹⁵.

L'automatisation conduit à une vision plus globale, dans laquelle les documents ne franchissent plus des étapes distinctes et successive, mais circulent dans des flux de traitement où la classification, le contrôle, l'extraction d'entités et les décisions de conservation peuvent intervenir simultanément¹⁹⁶. Ce glissement réactive la logique du *records continuum*, qui postule la présence simultanée des dimensions archivistiques (création, capture, organisation et plurilisation). L'IA, en intervenant sur plusieurs de ces dimensions en parallèle, renforce cette dynamique. Un même document peut en effet être décrit, enrichi, classé, évalué et rendu accessible dans un processus continu, en fonction des besoins des systèmes et des usagers¹⁹⁷.

d. L'archiviste comme garant d'intégrité, d'éthique et de fiabilité

L'intégration de l'IA dans la pratique archivistique renforce le rôle éthique de l'archiviste. Celui-ci devient un garant d'intégrité, responsable de la fiabilité des résultats produits par des systèmes potentiellement opaques ou biaisés¹⁹⁸. L'archiviste joue d'abord un rôle de médiateur entre les technologies et les usagers. Il doit être capable d'expliquer comment les résultats ont été obtenus, de contextualiser les extraits et de garantir que l'accès aux données respecte les droits, les règles de protection et la qualité requise¹⁹⁹. C'est particulièrement

¹⁹³ Echange par email avec Johan Pierret, délégué à l'archivage au Ministère de la Santé et de la Sécurité sociale du Grand-Duché du Luxembourg, 25 juillet 2025.

¹⁹⁴ Entretien avec Amandine Gorse, archiviste à la Chambre des Députés du Grand-Duché du Luxembourg, le 6 juin 2025, à l'International Archives Symposium, « Open data and AI new chances for archives ? » à Namur.

¹⁹⁵ BASMA Makhoul Shabou, « Archivistique à l'ère de l'IA : opportunités, défis et besoin d'utilisation responsable », dans *Arbido, Les archives du futur*, vol. 1, 2023, <https://arbido.ch/fr/edition-article/2023/archives-du-futur/archivistique-a-ler-de-lia-opportunites-defis-et-besoin-dutilisation-responsable>

¹⁹⁶ COLAVIZZA Giovanni e.a., « Archives and AI: An Overview of Current Debates and Future Perspectives », dans *Journal on Computing and Cultural Heritage*, vol.15, n°1, décembre 2021., p.2.

¹⁹⁷ Pour cette réflexion, l'outil *Copilot* a été utilisé en tant qu'aide à « l'exploration d'idées », sur la question de l'influence de l'IA sur le *records continuum*.

¹⁹⁸ SAAD Nora, ELMQADDEM Noureddine et VASSEUR Edouard, « L'intelligence artificielle au service de l'archivage numérique : optimisation des pratiques et défis éthiques », *op.cit.*, p.85 et MACLURE Jocelyn et SAINT-PIERRE Marie-Noëlle, *op.cit.*, p. 757.

¹⁹⁹ GAUTHIER Jean, « Portrait robot de l'archiviste de demain », dans *Archimag. Guide pratique n°81. Archivage physique et électronique : enjeux, méthodes et outil*, 2025, p.18-19.

essentiel dans un contexte où les modèles peuvent produire des hallucinations et des sorties erronées. Les centres d'archives sont ainsi régulièrement les témoins de demandes de consultation de fonds d'archives, qui ont été hallucinés par l'IA. L'archiviste intervient aussi dans la prévention de la désinformation, en s'assurant que les réponses générées, les résumés produits ou les rapprochements sémantiques ne déforment pas le sens des documents. La qualité des prompts, leur formulation et la maîtrise des limites du modèle deviennent ici des enjeux professionnels²⁰⁰. Enfin, l'archiviste doit assurer la médiation avec le public, qui tend parfois à faire un usage sans critique de l'IA et à croire à ses hallucinations, provoquant des quiproquos et une perte de confiance dans le monde archivistique, qui est pourtant essentiel en termes de gouvernance démocratique. Toutefois, dans des contextes institutionnels marqués par des logiques de rationalisation et d'accélération des traitements, il existe un risque que cette dimension critique soit progressivement réduite au profit de l'automatisation. L'usage de l'IA introduit en effet des systèmes fonctionnant sur des probabilités, des rapprochements statistiques et des résultats potentiellement contestables, ce qui implique d'accepter une certaine part d'incertitude et de risque dans les traitements documentaires. Or, la pratique archivistique repose historiquement sur des procédures stabilisées, normées et fortement encadrées. L'intégration de l'IA tend dès lors à déplacer la culture professionnelle vers une approche davantage fondée sur l'évaluation des risques, la supervision des traitements et la possibilité de contester ou de réévaluer les résultats produits par les systèmes.

Toutefois, l'intégration de l'IA dans les pratiques archivistiques soulève également des enjeux économiques et environnementaux. Le développement, l'entraînement et l'utilisation de certains modèles nécessitent des infrastructures techniques importantes, des capacités de calcul élevées et des ressources financières parfois considérables. À cela s'ajoutent des coûts liés au stockage, à la maintenance, aux abonnements et à la mise à jour des systèmes utilisés par les institutions. Sur le plan environnemental, les consommations énergétiques associées aux centres de données et aux traitements algorithmiques interrogent également la soutenabilité de ces technologies dans un contexte de multiplication des usages numériques. Dès lors, l'intégration de l'IA dans les services d'archives ne peut être pensée uniquement en termes d'efficacité technique, mais doit aussi prendre en compte les contraintes matérielles, économiques et écologiques qu'elle implique.

²⁰⁰ JOST Clémence, « Nicolas Curien : le rapport à la vérité est mal parti avec les IA génératives », dans *Archimag*, n°385, juin 2025, p.48-49.

Cette évolution implique également une transformation des compétences professionnelles. Il ne s'agit pas uniquement pour les archivistes de se former à l'usage de l'IA, mais aussi de développer une capacité critique à évaluer les outils disponibles, leurs limites, leurs coûts et leur pertinence selon les besoins réels des institutions et des traitements documentaires. Toutes les opérations ne nécessitent pas le recours à des modèles complexes ou particulièrement coûteux en ressources techniques, financières ou énergétiques. L'intégration de l'IA suppose dès lors de pouvoir sélectionner des outils adaptés aux objectifs poursuivis, dans une logique de proportionnalité, d'efficacité et de sobriété numérique.

Ainsi donc, l'IA ne supprime pas les pratiques professionnelles, elle les déplace, les recompose et les renforce. Les archivistes deviennent superviseurs, analystes, régulateurs et garants de la fiabilité et de l'éthique dans un environnement où l'automatisation rend plus visibles les risques et plus nécessaires les compétences humaines²⁰¹.

4. L'IA et la remise en tension du cycle de vie des archives

L'intégration de l'IA dans les environnements documentaires reconfigure la manière dont les documents sont créés, capturés, organisés et mis à disposition dans les organisations. L'IA n'agit pas à un moment unique de la trajectoire documentaire, mais intervient selon une logique simultanée, répétée et continue²⁰². Aussi, selon Giovanni Colavizza, le modèle du *record continuum* « *offers a holistic, pluralist framework covering all aspects and dimensions of archiving: the creation of records, their capturing, organizing, and pluralising. The model not only includes records, but also actors and processes that underlie and affect the creation, management, organization, and use of records* »²⁰³. Cette dynamique bouscule la linéarité du cycle de vie des archives, tout en renforçant la pertinence du *records continuum* comme cadre analytique. Elle ne provoque pas une rupture épistémologique, mais elle impose une lecture renouvelée de la temporalité documentaire, de la distribution des responsabilités et de la gouvernance des traitements²⁰⁴.

a. Une action simultanée qui bouscule la linéarité

Le cycle de vie repose traditionnellement sur la succession des âges courant, intermédiaire, et définitif, associés à des fonctions distinctes (qui sont, de façon caricaturale, la

²⁰¹ JACOB Steve, SOUISSI Seima et MARTINEAU Charlie, *op.cit.*, p.8.

²⁰² Cf. COLAVIZZA Giovanni e.a., *op.cit.*. Cet article propose une lecture des possibilités de l'IA dans l'archivistique, avec l'IA comme cadre d'analyse.

²⁰³ *Ibid.*, p.2.

²⁰⁴ *Ibid.*

production, l'usage administratif, puis la conservation ou la destruction). Cette séquentialité a longtemps structuré la discipline, car elle organise des responsabilités, des droits d'accès et des obligations légales.

Ces opérations se prolongent lors de la capture (OCR, HTR, enrichissement), se poursuivent pendant l'organisation (classification automatique, ajout de métadonnées, contrôle de cohérence) et influencent la pluralisation (recherche sémantique, agents conversationnels, valorisation). Elles peuvent de plus être relancées à n'importe quel moment, par le réentraînement des modèles, l'ajout d'un corpus, la révision d'une règle, etc²⁰⁵. Ce fonctionnement simultané et répété ne supprime pas les étapes, mais il rend plus poreuses les frontières temporelles du cycle de vie. L'enrichissement documentaire ne se limite plus aux premières phases du traitement, de même que la préparation du sort final n'intervient plus uniquement en aval. Les usages et les accès sur les documents influencent désormais leur organisation, et peuvent même réorienter certaines opérations de capture. La logique linéaire du cycle de vie devient dès lors difficile à mobiliser comme modèle descriptif des pratiques réelles, même si elle continue d'organiser les obligations institutionnelles. En effet, les traitements numériques sur les documents menés par l'IA ne s'inscrivent plus dans un ordre strict. Ils peuvent être mobilisés à tout moment, rejoués autant de fois que nécessaire, appliqués à des documents déjà conservés, ou intervenir en amont comme en aval des étapes théoriquement prévues. Ce n'est donc pas seulement la multiplication des opérations, que ce soit la classification, la détection, ou l'extraction, qui crée de la simultanéité, mais le fait que ces opérations deviennent modulaires, décomposables et détachées du « moment archivistique » auquel le cycle de vie prétend les assigner. Par exemple, un même document numérisé peut simultanément faire l'objet d'une reconnaissance d'écriture manuscrite, d'une extraction d'entités nommées, d'une génération automatique de métadonnées et d'une indexation thématique, sans que ces opérations correspondent à des étapes successives clairement séparées dans le temps.

b. Apports et limites du cycle de vie pour penser l'IA

Comme nous l'avons montré, la pratique quotidienne et les cadres réglementaires confirment que le cycle de vie conserve une grande pertinence. Il reste le modèle qui organise les obligations de conservation, la répartition des responsabilités, la gestion des durées et la

²⁰⁵ Cf. *Ibid.*

justification des éliminations. À ce titre, il demeure indispensable pour la gouvernance informationnelle et pour la preuve²⁰⁶.

Cependant, l'introduction de l'IA met en évidence plusieurs limites de ce modèle lorsqu'il s'agit de décrire ce qui se passe concrètement dans les environnements numériques. Trois phénomènes sont particulièrement significatifs. Le premier est celui des traitements répétés. Une classification peut être relancée, une extraction recalculée ou un enrichissement généré à nouveau au fil de l'évolution des modèles. Le second est celui des enrichissements successifs, qui produisent des couches dérivées (entités nommées, graphes, résumés, *embeddings*) sans correspondre à une étape précise du cycle. Le troisième est celui des transformations continues, où un même document peut être réévalué, recontextualisé ou vérifié à plusieurs moments, en fonction des besoins et des flux. Ces éléments montrent un décalage entre la linéarité normative du cycle de vie et la dynamique réelle des documents numériques lorsqu'ils sont traités par des systèmes automatisés d'IA. Ils soulèvent également la question du statut archivistique des différents outputs produits par ces traitements automatisés : doivent-ils être conservés, et si oui, selon quelles modalités et pour quelles finalités ?

c. Le *records continuum* comme cadre d'analyse renforcé

À l'inverse du cycle de vie, le *records continuum* est conçu pour appréhender des documents pris dans des flux continus où se combinent plusieurs dimensions : création, capture, organisation, pluralisation. Il semble que cette représentation est particulièrement adaptée aux effets de l'IA, car les opérations algorithmiques agissent précisément sur ces quatre dimensions en parallèle²⁰⁷. Le *continuum* permet aussi de penser la pluralité des agents impliqués, qu'ils soient producteurs, archivistes, systèmes d'information, modèles d'IA, ou utilisateurs finaux. L'IA introduit des agents semi-autonomes capables d'extraire, d'indexer ou de proposer des décisions, de rétention, de typage, ou encore de rédaction, ce qui redistribue l'agentivité entre humains et machines. Dans cette configuration, l'archiviste n'est pas dépossédé, il devient un coordinateur du sens et un superviseur des opérations²⁰⁸.

Cette perspective introduit un élément déterminant : celui de la nécessité de pouvoir expliquer et documenter les transformations opérées par les systèmes. Les *paradata* prennent ici une portée théorique particulière. Elles permettent de rendre visibles les enchaînements

²⁰⁶ GUYON Céline, « Le temps long de l'archive », *Séminaire de l'Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB)*, juin 2019, [tps://hal.science/hal-04371293v1](https://hal.science/hal-04371293v1), consulté le 19 juin 2025.

²⁰⁷ COLAVIZZA Giovanni e.a., *op.cit.*, p.2.

²⁰⁸ JACOB Steve, SOUISSI Seima et MARTINEAU Charlie, *op.cit.*, p.8.

d'actions, les paramètres mobilisés et les agents impliqués, et constituent ainsi un moyen de donner forme et lisibilité à la simultanéité propre au *continuum*.

Conclusions

Les transformations liées au numérique et celles introduites par l'IA s'inscrivent dans un ensemble d'évolutions qui ont modifié la manière de produire, de gérer et d'utiliser les documents. Le numérique a entraîné une reconfiguration de la nature documentaire et des conditions pratiques de production, de gestion et de conservation. Il a conduit à redéfinir la fixité, désormais fondée sur l'intégrité plutôt que sur le support, et à renforcer l'usage des métadonnées, la maîtrise des flux et une gestion plus structurée des processus documentaires. Ces évolutions ont aussi rendu moins nette la séparation entre les étapes du cycle de vie, dans des systèmes où plusieurs opérations sont désormais regroupées.

L'IA intervient dans un autre registre. Elle agit sur les contenus, leur structure et leur traitement, en produisant des entités nommées, des segments, des résumés ou des classifications qui complètent les documents existants. Ces dérivés facilitent l'accès et les rapprochements entre documents, tout en introduisant une couche d'interprétation automatisée qui doit être distinguée des métadonnées archivistiques. Leur identification repose notamment sur la documentation des traitements, via les *paradata*, qui rendent visible la succession des opérations et soutiennent des logiques *d'accountability* dans des environnements où certains processus demeurent opaques. L'IA peut aussi intervenir dans les dispositifs de gestion et de conservation, en accélérant ou en orientant certaines décisions.

La comparaison des deux dynamiques met en évidence des similitudes et une continuité entre elles : dans les deux cas, les frontières entre production, gestion, description et conservation deviennent plus souples, ce qui invite à envisager les processus documentaires selon une logique plus continue. Si le numérique avait déjà atténué le caractère séquentiel du cycle de vie, l'IA prolonge ce mouvement en intervenant au moment même où les documents sont produits ou mobilisés. Ces évolutions peuvent être éclairées par des approches inspirées du *records continuum*, qui envisagent les processus comme simultanés et ouverts à des usages multiples.

Comme le numérique, l'IA conduit à repenser le modèle archivistique traditionnel fondé sur des étapes nettement séparées, sans remettre en cause les principes fondamentaux de la discipline. Des différences apparaissent toutefois puisque le numérique constituait avant tout un changement technologique et documentaire, tandis que l'IA introduit en plus un changement décisionnel, en influençant la manière dont les documents sont catégorisés, interprétés ou signalés. Elle participe ainsi à la gouvernance des archives, en déplaçant certaines opérations vers des traitements automatisés qui doivent être explicités, documentés et supervisés.

Ces éléments conduisent à considérer que l'IA ne constitue peut-être pas une rupture de même nature que celle qu'a représentée la dématérialisation du numérique, mais qu'elle semble plutôt amplifier et reconfigurer certaines tendances déjà présentes. Toutefois, elle introduit également de nouvelles transformations, notamment par l'automatisation de certaines opérations d'analyse, de classement, d'extraction ou de production d'information. Elle complexifie ainsi les conditions de mise en œuvre des principes archivistiques, en particulier en matière de provenance, d'authenticité, de fixité et de gestion des représentations. Dans ce contexte, les modèles du cycle de vie et du continuum demeurent pertinents, mais doivent intégrer des opérations plus nombreuses, plus distribuées et parfois moins explicites.

Enfin, ces transformations redéfinissent la place du professionnel et son agentivité. La supervision des traitements, la documentation des opérations, l'analyse des résultats produits automatiquement et la compréhension des logiques de modélisation deviennent des compétences essentielles. Dans un environnement où les opérations se diversifient, la maîtrise des *paradata*, l'attention portée à la transparence et la capacité à situer les dérivés dans la chaîne documentaire constituent des conditions nécessaires pour préserver la cohérence et la fiabilité des archives.

Sources et bibliographies

« AI Key Terms », sur *Chapman University*, <https://www.chapman.edu/ai/key-terms.aspx>

« Apprentissage profond : la mécanique de la magie », sur *ISO*, <https://www.iso.org/fr/intelligence-artificielle/apprentissage-profond-deep-learning>

« ChatGPT, un beau parleur bien entraîné », dans *LINC, Laboratoire d'Innovation Numérique de la CNIL*, <https://linc.cnil.fr/dossier-ia-generative-chatgpt-un-beau-parleur-bien-entraine>

« Comment une IA générative crée-t-elle du texte ? », sur *Délégation Régionale Académique Numérique Educatif*, <https://drane-versailles.region-academique-idf.fr/spip.php?article845>

Commission Nationale de l'Informatique et des Libertés (CNIL), <https://www.cnil.fr/>

« Conserver de l'information », sur *Archives Générales du Royaume*, <https://www.arch.be/index.php?l=fr&m=fonctionnaire&r=conserver-de-l-information>

DataFranca, *Les 101 mots de l'intelligence artificielle. Petit guide du vocabulaire essentiel de la science des données et de l'intelligence artificielle*, <https://datafranca.org/>

« Différence entre un système d'IA et un modèle d'IA », sur *Dastra*, <https://www.dastra.eu/fr/article/difference-entre-un-systeme-dia-et-un-modele-dia/57625>

EU Artificial Intelligence Act, <https://artificialintelligenceact.eu/fr/>

« Intelligence artificielle : définition et utilisation », sur *Parlement européen*, <https://www.europarl.europa.eu/topics/fr/article/20200827STO85804/intelligence-artificielle-definition-et-utilisation>

« Intelligence artificielle et collections de la BnF : l'exemple de l'HTR (Handwritten Text Recognition) », sur *Bibliothèque nationale de France*, <https://www.bnf.fr/fr/intelligence-artificielle-et-collections-de-la-bnf-lexemple-de-lhtr-handwritten-text-recognition>

InterPARES Trust AI, <https://interparestrustai.org/>

« Journée internationale de la gouvernance de l'information – interview d'Amandine Gorse », sur *Labgroup*, <https://www.labgroup.com/journee-internationale-de-la-gouvernance-de-linformation-interview-damandine-gorse/>

« Les secrets du traitement du langage naturel décryptés », sur *ISO*, <https://www.iso.org/fr/intelligence-artificielle/traitement-langage-naturel>

« Qu'est-ce que l'intelligence artificielle (IA) », sur *ISO*, <https://www.iso.org/fr/contents/news/insights/AI/what-is-ai-all-you-need-to-know.html>

« Qu'est-ce qu'un modèle IA ? », sur *IBM*, <https://www.ibm.com/fr-fr/think/topics/ai-model>

« Qu'est-ce qu'un système d'intelligence artificielle ? », sur *Autorité de protection des données*, <https://www.autoriteprotectiondonnees.be/citoyen/themes/intelligence-artificielle/-qu-est-ce-qu-un-systeme-d-intelligence-artificielle->

« Tout connaître de l'apprentissage machine (ML) », sur *ISO*, <https://www.iso.org/fr/intelligence-artificielle/apprentissage-machine>

ALAMMAR Jay, « The illustrated Word2vec », <https://jalanmar.github.io/illustrated-word2vec/>

ALAOUI Siham, « L'intelligence artificielle et la gestion documentaire : quels apports ? Quels enjeux ? », dans *La Revue canadienne des sciences de l'information et de bibliothéconomie*, vol. 46, n°2, 2023, p. 1-20.

AZAD Reza e.a., « Advances in medical image analysis with Transformers : a comprehensive review », dans *Medical Image Analysis*, vol 91, 2024.

BANAT-BERGER Françoise, « Les fonctions de l'archivistique à l'ère du numérique », dans DELPIERRE Nicolas, HIRAUX Françoise et MIRGUET Françoise (éd.), *Les chantiers du numérique. Dématérialisation des archives et métiers de l'archiviste*, Louvain-la-Neuve, L'Harmattan, 2012, p.39-59.

BANAT-BERGER Françoise, « Les transformations du métier de l'archiviste », dans SERVAIS Paul et MIRGUET Françoise (éd.), *Archivistes de 2030. Réflexions prospectives*, Louvain-la-Neuve, L'Harmattan, 2011, p.33-45.

BANAT-BERGER Françoise, « Un métier à part entière, l'archiviste un généraliste de l'information : qu'en est-il en 2012 dans le nouvel environnement numérique ? », dans *La Gazette des archives*, n° 226, vol.2, 2012, p.117-126.

BASMA Makhoul Shabou, « Archivistique à l'ère de l'IA : opportunités, défis et besoin d'utilisation responsable », dans *Arbido, Les archives du futur*, vol. 1, 2023, <https://arbido.ch/fr/edition-article/2023/archives-du-futur/archivistique-a-ler-de-lia-opportunites-defis-et-besoin-dutilisation-responsable>

BASMA Makhoul Shabou, « Fonction d'évaluation des archives. Bilan sommaire des développements, des enjeux actuels et des défis futurs », dans GAGNON-ARGUIN Louise et

LAJEUNESSE Marcel (éd.), *Panorama de l'archivistique contemporaine. Évolution de la discipline et de la profession. Mélanges offerts à Carol Couture*, Québec, Presses de l'Université du Québec, 2015, p.195-218.

BASMA Makhoul Shabou, « Humanités numériques & Sciences de l'information : ressemblance, complémentarité et développements mutualisés », dans *Journal of information sciences*, vol.1, n°21, 2022, p.18-30.

BRULEAUX Anne-Marie, « Former l'archiviste du XXI^e siècle », dans *La Gazette des archives*, vol.226, n°2, 2012, p. 71-93.

BURNEL Anne et SERVANT Hélène, « Vers une nouvelle pratique archivistique », dans *La Gazette des Archives*, n°223, vol.3, 2011, p. 71-84.

CASTANIÉ Camille, « L'intelligence artificielle permet de nouveaux projets pour les archivistes », dans *Archivistes !*, n°154, octobre-décembre 2025, p.21-22.

CAYA Marcel, « La théorie des trois âges en archivistique. En avons-nous toujours besoin ? », conférence donnée à l'Ecole nationale des chartes, le 2 décembre 2004, <http://elec.enc.sorbonne.fr/conferences/caya>

CHABIN Marie-Anne, « L'archiviste de 2030 entre archives numériques et utilisateurs connectés », dans SERVAIS Paul et MIRGUET Françoise (éd.), *Archivistes de 2030. Réflexions prospectives*, Louvain-la-Neuve, L'Harmattan, 2015, p.61-72.

CHABIN Marie-Anne, « Le Records management : concepts et usages », 2012, <https://www.marieannechabin.fr/arcateg/wp-content/uploads/2017/03/MAC-Le-Records-management.-Concept-et-usages-2012.pdf>

CHABIN Marie-Anne, « Les deux âges des archives », sur *Le blog de Marie-Anne Chabin*, novembre 2022, <https://www.marieannechabin.fr/2022/11/les-deux-ages-des-archives>

COLAVIZZA Giovanni e.a., « Archives and AI: An Overview of Current Debates and Future Perspectives », dans *Journal on Computing and Cultural Heritage*, vol.15, n°1, décembre 2021.

COUTURE Carol, « La discipline archivistique au Canada : état de développement et perspectives d'avenir », sur *In Situ*, n°30, 2016, <http://journals.openedition.org/insitu/13669>

COUTURE Carol, « Le cycle de vie des documents d'archives », dans ROUSSEAU Jean-Yves et COUTURE Carol (éd.), *Les fondements de la discipline archivistique*, Québec, Presses universitaires du Québec, 2008, p.95-114.

DAVET Jeremy, HAMIDZADEH Babak, FRANKS Patricia, « Archivist in the machine: paradata for AI-based automation in the archives », dans *Archival Science*, vol.23, 2023, p.275-295.

DEBEAUVAIS Jeanne, « Intégrer l'IA dans la gestion documentaire : un retour d'expérience », dans *Archivistes !*, n°155, janvier-mars 2026, p.26.

DEMOULIN Marie, « Quelques aspects juridiques de l'archivage électronique », dans DELPIERRE Nicolas, HIRAUX Françoise et MIRGUET Françoise (éd.), *Les chantiers du numérique. Dématérialisation des archives et métiers de l'archiviste*, Louvain-la-Neuve, L'Harmattan, 2012, p.21-38.

DURANTI Luciana, « The Archival Bond », dans *Archives and Museum Informatics*, vol.11, n°3, septembre 1991, p.213-218.

DURANTI Luciana et McNEIL Heather, « The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project », dans *Archivaria*, n°42, 2001, p.46-67.

DURANTI Luciana et ROGERS Corinne, « Artificial Intelligence and Documentary Heritage », Memory of the World Sub on Education and Research (SCEaR), Newsletter 2024. Special Issue 2024.

DURANTI Luciana et THIBODEAU Ken, « The Concept if Record in Interactiven Experiential and Dynamic Environments : the View of InterPARES », dans *Archives and Museum Informatics*, vol.6, n°1, mars 2006, p.13-68.

EL MEKKI Touria Ait et GRAILLES Bénédicte, « Sélectionner, épurer, décrire. L'apport de l'IA dans le traitement archivistique des messages électroniques », dans *Culture et recherche*, 2024, vol.147, p.43-45.

FELICIATI Pierluigi et DURANTI Luciana, « The responsible use of artificial intelligence in archives through the use of paradata », dans *JLIST.it*, vol.6, n°2, mai 2025.

FLYNN Sarah J. A., « The records continuum model in context and its implications for archival practice », dans *Journal of the Society of Archivists*, vol.22, n°1, p.79-93.

FRANKS Patricia, « In the pursuit of archival accountability : positioning paradata as AI processual documentation », *Society of American Archivists, Research Forum 2023*, https://www2.archivists.org/sites/all/files/Franks_In%20the%20Pursuit%20of%20Archival%20Accountability.pdf

FRINGS-HESSAMI Viviane, « La perspective du Continuum des archives illustré d'un document personnel », dans *Revue électronique suisse de science de l'information*, n°149, 2018.

FUENTES-HASHIMOTO Lourdes, « L'archivage électronique à l'épreuve de la théorie des trois âges » sur *Archives Online*, <https://archivesonline.wordpress.com/2010/08/30/larchivage-electronique-a-lepreuve-de-la-theorie-des-trois-ages/>

FUENTES-HASHIMOTO Lourdes, VERDO Rémy et WATEL François, « La tradition de records management aux Archives diplomatiques et son adaptation à l'environnement électronique », dans *Quand l'archivage devient électronique. Actes du colloque international organisé par les Archives diplomatiques et les Archives nationales, 5 et 6 février 2013*, p.25-40.

GAUTHIER Jean, « Portrait-robot de l'archiviste de demain », dans *Archimag. Guide pratique n°81. Archivage physique et électronique : enjeux, méthodes et outil*, 2025, p.18-19.

GRAILLES Bénédicte et AIT EL MEKKI Touria, « De la boîte noire à la boîte à surprise, le programme Pôle-mél et les messageries électroniques », Journée d'étude AAF-Ecole des Chartes, L'intelligence artificielle dans la gestion des données, 4 juillet 2023.

GRINBAUM Alexei e.a., « Systèmes d'intelligence artificielle générative : enjeux d'éthique ». Comité national pilote d'éthique du numérique, 2023, <https://cea.hal.science/cea-04153216/>, consulté le 6 juin 2025.

GUYON Céline, « L'archivage comme dispositif de transformation de la nature intrinsèque des objets nativement numériques », dans *Balisages*, n°1, 2020, <http://journals.openedition.org/balisages/282>.

GUYON Céline, « Le temps long de l'archive », *Séminaire de l'Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB)*, juin 2019, <https://hal.science/hal-04371293v1>

GUYON Céline, « Théorie, technique et pratique archivistique en environnement numérique. Colloque International sur le Document Electronique : Document et archivage : pratiques formelles et informelles dans les organisations », *CIDE23*, 2023, Grenoble, France, sur <https://hal.science/hal-04371177v1>

GUYON Céline, « Une archivistique sous influence. La pratique archivistique face aux innovations technologiques (1960-2022) », dans *Revue d'Histoire culturelle, XVIIIe-XXIe siècles*, n°5, 2022, <https://journals.openedition.org/rhc/3466>

HENNO Jacques, « 1956 : et l'intelligence artificielle devint une science », sur *Les Echos*, <https://www.lesechos.fr/2017/08/1956-et-lintelligence-artificielle-devint-une-science-181042>

JACOB Steve, SOUISSI Seima et MARTINEAU Charlie, *Intelligence artificielle et transformation des métiers de la gestion documentaire*, Chaire de recherche sur l'administration publique à l'ère numérique, Université de Laval, 2022, https://bibliopiaf.ebsi.umontreal.ca/bibliographie/8WWN5HAY/download/SDIBUXKJ/ia_et_metiers_de_la_gestion_documentaire.pdf

JOST Clémence, « Le baromètre des usages des IA génératives par les métiers de l'information », dans *Archimag. Guide pratique n°83. IA génératives : cas d'usage et retours d'expérience*, février 2025, p.4-9.

JOST Clémence, « Nicolas Curien : le rapport à la vérité est mal parti avec les IA génératives », dans *Archimag*, n°385, juin 2025, p.48-49.

KAËLIG Alléaume, « Data librarian, au coeur du cycle de vie des données de la recherche », dans *Archimag*, n°387, septembre 2025, p.18.

KAËLIG Alléaume, « SNCF : un chatbot embarque les visiteurs sur les rails du patrimoine », dans *Archimag*, n°380, décembre 2024-janvier 2025, p.20.

KERN Gilliane, HOLGADO Sandra, COTTIN Michel, « Cinquante nuances de cycle de vie. Quelles évolutions possibles ? », dans *Les cahiers du numérique*, n°2, 2015, p.37-76.

KLEIN Anne, *Archive(s), mémoire, art. Éléments pour une archivistique critique*, Laval, Presses universitaires de Laval, 2019.

LEMAY Yvon et KLEIN Anne, « Les archives définitives : un début de parcours. Revisiter le cycle de vie et le *Records continuum* », dans *Archivaria*, n°77, 2014, p.73-102.

LORRE Benjamin, *L'intelligence artificielle générative. Explications, modèles, enjeux*, Paris, ISTE Editions, 2025.

MACLURE Jocelyn et SAINT-PIERRE Marie-Noëlle, « Le nouvel âge de l'intelligence artificielle : une synthèse des enjeux éthiques », dans *Les cahiers de propriété intellectuelle*, vol. 30, n°3, 2018, p. 741-765.

MCKEMMISH Sue e.a. (éd.), *Archives : Recordkeeping in Society*, New South Wales, Centre for Information Studies, 2005.

MITTAL Aayush, « NLP Rise with Transformer Models | A Comprehensive Analysis of T5, BERT, and GPT », sur *Unite.ai*, <https://www.unite.ai/nlp-rise-with-transformer-models-a-comprehensive-analysis-of-t5-bert-and-gpt/>

MUEHLBERGER Guenter e.a., « Transforming scholarship on the archives through handwritten text recognition. Transkribus as a case study », dans *Journal of Documentation*, vol 75, n°5, 2029, p. 954-976.

NAHUET Robert, « L’archivistique contemporaine à l’âge adulte : pertinence et actualité du respect des fonds », dans *Archives*, vol.41, n°1, 2009-2010, p.45-60.

NIEUWBOURG Philippe, « La gouvernance de l’IA, entre conformité et bonnes pratiques », sur *Archimag*, 11 décembre 2024, <https://www.archimag.com/univers-data/2024/12/11/gouvernance-intelligence-artificielle-conformite-bonnes-pratiques>.

PÉROTIN Yves, « L’administration et les « trois âges » des archives », dans *Seines et Paris*, n°20, octobre 1961, p.31-33.

POIRIER Charles-Antoine, *AI. Termes et concepts*, Editions Futur Path, 2025.

RAJOTTE David, « La réflexion archivistique à l’ère du document numérique : un bilan historique », dans *Archives*, vol.42, n°2, 2010-2011, p.69-105.

REMIZE Michel, « Automatisation : aller encore plus loin avec l’IA générative », dans *Archimag*, n°380, décembre 2024-janvier 2025, p.26-27.

ROUSSEAU Jean-Yves et COUTURE Carol, « La place de l’archivistique dans la gestion de l’information », dans ROUSSEAU Jean-Yves et COUTURE Carol (éd.), *Les fondements de la discipline archivistique*, Québec, Presses universitaires du Québec, 2008, p.42-57.

SAAD Nora, ELMQADDEM Nouredine, VASSEUR Edouard, « Archivage à l’ère du numérique : objet et cycle de vie », dans *Journal of Information Sciences*, vol.21, n°2, 2022, p.173-193.

SAAD Nora, ELMQADDEM Nouredine et VASSEUR Edouard, « L’intelligence artificielle au service de l’archivage numérique : optimisation des pratiques et défis éthiques », dans *Journal of Information Sciences*, vol.3, n°1, 2024, p.68-89.

Services des ressources informationnelles et archives de l’Université de Lausanne, *Politique de records management. Glossaire*, 2014, https://www.unil.ch/files/live/sites/unil/files/02-universite/0204-organisation/020404-unites-et-services/UNIRIS/references/UNIL_POL_RM_Glossaire_VF.pdf

Service des ressources informationnelles et archives de l'Université de Lausanne, *Politique de records management. Records management et archivage électronique : cadre normatif*, 2014, https://www.unil.ch/files/live/sites/unil/files/02-universite/0204-organisation/020404-unites-et-services/UNIRIS/unites/UNIL_POL_RM_Archives_electro_cadre_normatif_VF.pdf.

Services des ressources informationnelles et archives de l'Université de Lausanne, « Records management et gouvernance informationnelle : Concepts et explications », 2016, https://www.unil.ch/files/live/sites/unil/files/02-universite/0204-organisation/020404-unites-et-services/UNIRIS/references/UNIL_RM_Gouvernance_Info_concepts_explications_VF-1.pdf

SIVAGAMI Casimir, « Archiviste : enquête sur un métier en constante évolution », dans *Archimag*, n°391, février 2026, p.22-23.

SIVAGAMI Casimir, « Pêle-mél pour l'archivage des e-mails », dans *Archimag. Guide pratique n°83. IA générative : cas d'usage et retours d'expérience*, p.40.

SIVAGAMI Casimir, « Stéphane Wulc : la complexité comme terrain de jeu », dans *Archimag*, n°380, décembre 2024-janvier 2025, p.43.

Techopedia, <https://www.techopedia.com/>

TEXIER Bruno, « IA et archives : du patrimoine culturel aux documents d'entreprise », dans *Archimag. Guide pratique n°77. L'IA générative : l'utiliser dans ses projets professionnels*, juin 2024, p.26-27.

TEXIER Bruno, « L'archivage électronique au temps d'eiDAS 2 », dans *Archimag*, n°387, septembre 2025, p.24.

TEXIER Bruno, « L'IA en mode majeur à la Philharmonie de Paris », dans *Archimag*, n°380, décembre 2024-janvier 2025, p.19.

TEXIER Bruno, « Université de Montréal : l'IA pour déchiffrer des documents manuscrits », dans *Archimag*, n°380, décembre 2024-janvier 2025, p.16.

VELLUET Hugo, « IA générative et gestion documentaires : comment ça marche ? », dans *Archimag. Guide pratique*, n°77, p.38-39.

VÉROVE Simon et BAHIER MICHEL Livia, « Réflexion sur les apports de l'IA au métier. IA : vers une réalité archiviste ? », dans *Archivistes !*, n°47, janvier-mars 2024, p.33-34.

UPWARD Franck, « Structuring the records continuum, part one : postcustodial principles and properties », dans *Archives and Manuscripts*, vol.24, n°2, 1996, p. 268-285.

WU Xingjiao, « A survey of human-in-the-loop for machine learning », dans *Future Generation Computer System*, vol. 135, 2022, p. 363-381.

ZHANG Alexander et VARSHNEY Lav R., « Conceptualizing Agency : A Framework for Human-AI Interaction », dans *CEUR Workshop Proceedings*, n°3957, 2025, <https://experts.illinois.edu/en/publications/conceptualizing-agency-a-framework-for-human-ai-interaction/>

Table des matières

Introduction	1
Prolégomènes. Intelligence artificielle : définitions et fonctionnement	6
1. Définitions de base	6
2. Comment fonctionne l'IA ?.....	12
a. Modèle génératif	13
b. Modèle discriminatif.....	15
c. Traitement d'images	16
3. Conclusion.....	17
Première partie. La pratique archivistique face aux enjeux du numérique avancé.	18
1. Le numérique et la forme documentaire. Redéfinir l'authenticité, l'intégrité et le lien archivistique à l'ère numérique	19
2. Le numérique et les dispositifs de gestion et de conservation. Garantir la valeur probante et la mémoire à long terme des archives numériques.....	21
3. Le numérique et l'évolution des pratiques professionnelles. De la conservation à la gouvernance informationnelle : transformations du métier d'archiviste à l'ère numérique	23
4. Le numérique et la mise à l'épreuve du cycle de vie des archives. Le numérique et la remise en question du cycle de vie des archives	25
5. Conclusion.....	30
Deuxième partie. La pratique archivistique face aux enjeux de l'intelligence artificielle	32
1. L'IA et la forme documentaire.....	35
a. Un accès renouvelé aux contenus textuels et manuscrits.....	36
b. Extraction d'information et enrichissement automatique des métadonnées	37
c. Une structuration computationnelle du document : vers une forme documentaire augmentée.....	38
d. Conditions, biais et limites.....	39

e.	Limites théoriques : métadonnées, authenticité et statut du record à l'ère de l'automatisation	40
2.	L'IA et les dispositifs de gestion et de conservation.....	41
a.	Automatisation des processus documentaires.....	41
b.	IA et conservation : rétention, contrôle, conformité	42
c.	Recherche enrichie et nouvelles formes de valorisation.....	43
3.	L'IA et les pratiques professionnelles	44
a.	Évolution des gestes : de l'exécution au pilotage	44
b.	Nouvelles compétences et collaborations étendues	45
c.	L'IA comme intensification du mouvement vers le <i>records management</i>	45
d.	L'archiviste comme garant d'intégrité, d'éthique et de fiabilité.....	46
4.	L'IA et la remise en tension du cycle de vie des archives	48
a.	Une action simultanée qui bouscule la linéarité.....	48
b.	Apports et limites du cycle de vie pour penser l'IA	49
c.	Le <i>records continuum</i> comme cadre d'analyse renforcé	50
	Conclusions	52
	Sources et bibliographies.....	54

