

Vers l'intelligence artificielle et au-delà ! Une feuille de route pour la BnF

En 2020, la Bibliothèque nationale de France s'est dotée d'une feuille de route sur l'intelligence artificielle afin de mieux répondre aux nombreux défis, technologiques, professionnels, culturels, éthiques posés par l'IA.

{ BnF

Depuis plusieurs années déjà, la Bibliothèque nationale de France (BnF) travaille avec des partenaires du secteur académique sur des expérimentations à base d'intelligence artificielle (IA), notamment dans le domaine de l'OCR et de l'analyse d'images (computer vision). Elle a mis en place avec l'IR Huma-Num le BnF DataLab, qui vise à développer de nouveaux usages de recherche sur les collections numériques massives, mobilisant entre autres la fouille de données et l'apprentissage machine. Cependant, si ces projets expérimentaux portant sur des corpus réduits montrent des résultats satisfaisants, les industrialiser à l'échelle des 9 millions de documents de Gallica est une autre affaire. Plus généralement, l'intelligence artificielle représente un défi à bien des égards pour une institution comme la BnF : parce que le terme est devenu omniprésent sans pour autant être clairement défini, parce que la technologie montre une grande maturité dans le privé mais que les cas d'usage propres aux bibliothèques sont encore en devenir, parce qu'elle représente des investissements considérables et remet en cause l'infrastructure,

la gouvernance des données, l'organisation du travail; enfin, parce qu'elle pose de nombreuses questions éthiques quant à son impact sur l'environnement et plus encore, sur l'humain.

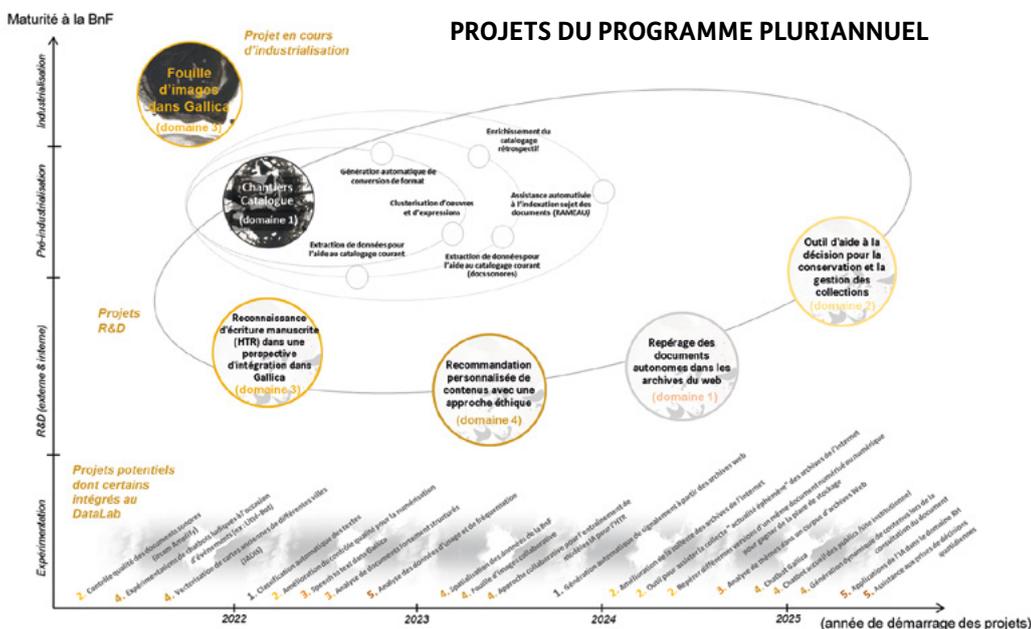
En 2018, la première conférence « Fantastic Futures » organisée par la Bibliothèque nationale de Norvège et la bibliothèque de Stanford nous a convaincus de l'importance de nous doter d'une feuille de route pour nous mettre en ordre de marche et accueillir le potentiel transformateur de l'IA. Le besoin s'est donc fait sentir de formaliser la stratégie de l'établissement dans ce domaine afin d'éclairer la route pour les années à venir : c'était l'objectif de la lettre de mission confiée en septembre 2020 par le directeur général de l'établissement à Emmanuelle Bermès, premier signe d'une volonté institutionnelle d'affirmer cette orientation nouvelle.

CINQ ACTIONS POUR SE LANCER

Élaborée au moyen d'une série d'ateliers réunissant les collègues intéressés, d'une enquête interne, d'un parangonnage international et d'une veille extensive, la feuille

de route synthétise le chemin à parcourir en cinq actions.

1. La première action à entreprendre consiste à inscrire l'IA dans la stratégie de l'établissement. Alors que se posait justement la question du renouvellement du contrat d'objectifs et de performance avec la tutelle pour la période 2022-2027, cette opportunité a été saisie pour faire de l'IA une dynamique transverse du nouveau contrat, susceptible d'apporter des pistes d'innovation dans tous les domaines métier de la bibliothèque : des entrées numériques à la médiation, du signalement à la conservation, du traitement des documents numérisés à leur accès dans Gallica, de l'analyse des usages des publics aux fonctions de gestion administrative. Cette stratégie globale met l'accent, dès cette étape, sur les enjeux éthiques, en particulier la question des données personnelles, les risques liés aux biais dans les données, l'impact environnemental de cette technologie et l'accompagnement du changement. Celui-ci touche particulièrement la fonction informatique, qui doit se mobiliser pour devenir une force motrice des évolutions à venir.



Ces projets relèvent de 5 domaines

1. Aide au catalogage et signalement
2. Gestion des collections, des entrées à la conservation
3. Exploration, analyse des collections et amélioration de l'accès
4. Médiation, valorisation et éditorialisation des collections
5. Aide à la décision et au pilotage

Information et contact
www.bnf.fr/fr/feuille-de-route-ia

2. La deuxième action se focalise sur les projets de recherche et développement, ainsi que sur les mesures à prendre pour faciliter l'industrialisation des résultats lorsque ceux-ci sont convaincants.

3. La troisième action porte sur le développement des compétences, non seulement en matière d'expertise pour les agents (informaticiens ou bibliothécaires) qui auront à participer aux projets mobilisant l'intelligence artificielle, mais aussi pour l'ensemble du personnel. En effet, pour les professionnels des bibliothèques, comprendre ce qu'est l'intelligence artificielle, comment elle fonctionne et les risques qu'elle présente relève de la culture générale. Au-delà de la stratégie de la BnF sur ce sujet, il s'agit d'une question de société qui rejoint notre mission de sensibilisation des publics. Les algorithmes sont présents dans notre quotidien, ils utilisent nos données et influencent nos décisions : avoir conscience de ces mécanismes est un enjeu citoyen.

4. La quatrième action est la plus technique : industrialiser l'IA nécessite d'adapter l'infrastructure informatique, mais aussi d'agir sur la gestion des données et leur qualité, un point fort de la BnF de par sa mission mais qui se heurte encore trop, dans le système d'information actuel, au silotage des collections de nature différente (par exemple, les archives Web d'une part et la bibliothèque numérique d'autre part).

5. Enfin, la cinquième action vise à doter la BnF d'un programme pluriannuel, avec des partenaires du secteur académique mais aussi du privé, et d'autres bibliothèques, afin de créer du lien entre les projets et de mutualiser les briques qui peuvent l'être.

UNE GALAXIE DE PROJETS EN DEVENIR

Dans la feuille de route, le programme pluriannuel prend la forme d'une galaxie de projets, représentés ici en fonction de leur niveau de priorité, de la maturité de la BnF sur le cas d'usage concerné et d'une ébauche de planification. Chaque planète représente un projet considéré, durant l'étude, comme majeur.

En haut à gauche, le projet le plus immédiat mais aussi le plus mature est l'industrialisation de la fouille d'images dans Gallica, qui fait suite à plus de dix ans d'expérimentations et de R&D sur ce sujet. Financé par France Relance dans le cadre de l'appel à projet « Numérisation de l'architecture et du patrimoine », en partenariat avec l'INHA (Institut

national d'histoire de l'art) et la bibliothèque nationale et universitaire de Strasbourg, le projet Gallica Images sera lancé en 2023 ; il utilisera l'IA pour segmenter et caractériser plusieurs millions de contenus iconographiques aujourd'hui peu accessibles dans Gallica.

Ensuite viennent d'autres projets jugés essentiels pour le développement de la BnF dans les années à venir : une planète dotée de nombreux satellites symbolise les expérimentations diverses à mener dans le champ des catalogues ; la reconnaissance d'écritures manuscrites (HTR) et la recommandation personnalisée avec une approche éthique sont des sujets sur lesquels la coopération avec d'autres institutions s'annonce prometteuse. Enfin, dans le champ de la conservation, outiller le futur site d'Amiens d'une intelligence de la donnée est l'une de nos perspectives de long terme, une ambition qui nécessite de poursuivre les expérimentations, telles que le projet Dalgocol récemment achevé.

Au bas du schéma, dans une zone métaphoriquement encore plongée dans des brumes d'incertitudes, sont recensés tous les autres cas d'usage qui ont émergé lors de l'élaboration de la feuille de route, mais qui n'ont pas été, à ce jour, priorités.

Le BnF DataLab sera sans nul doute un dispositif essentiel pour leur développement.

ET APRÈS 2026 ?

Si la BnF s'est dotée, avec cette feuille de route, de jalons à atteindre et livrables à réaliser d'ici à 2026, il est certain qu'on ne pourra pas parler alors d'achèvement. Il s'agit davantage d'enclencher une dynamique : l'intelligence artificielle trouve dans les bibliothèques un terrain de développement naturel, à la croisée des humanités et des technologies, mais n'est pas pour autant une fin en soi. Élément d'une culture numérique plus vaste, elle nous amène à nous interroger sur les principes de transparence, d'explicabilité, de justice (équité/égalité) et de sobriété qui devraient régir toute innovation numérique, et qui pourront sans nul doute nous guider dans les nouveaux défis qui émergeront à l'avenir, au-delà de l'IA.

EMMANUELLE BERMÈS

Adjointe pour les questions scientifiques et techniques auprès du directeur des services et des réseaux, Bibliothèque nationale de France
emmanuelle.bermes@bnf.fr

CÉLINE LECLAIRE

Chargée de production de supports stratégiques, Bibliothèque nationale de France
celine.leclaire@bnf.fr

● ● ● REPÈRES BIBLIOGRAPHIQUES

La BnF et l'intelligence artificielle, feuille de route :

www.bnf.fr/fr/feuille-de-route-ia

« L'intelligence artificielle à la BnF », Dossier Grand angle, *Chroniques* n° 93, janvier-mars 2022. http://chroniques.bnf.fr/pdf/Chroniques_93.pdf

Jean-Philippe Moreux, « Recherche d'images dans les bibliothèques numériques patrimoniales et expérimentation de techniques d'apprentissage profond », *Documentation et bibliothèques*, volume 65, numéro 2, avril-juin 2019, p. 5-27. <https://id.erudit.org/iderudit/1063786ar>

Emmanuelle Bermès, Eleonora Moiraghi, « Le patrimoine numérique national à l'heure de l'intelligence artificielle. Le programme de recherche Corpus comme espace d'expérimentation pour les humanités numériques », *Revue Ouverte d'Intelligence Artificielle*, Volume 1 (2020) no. 1, pp. 89-109. <https://roia.centre-mersenne.org/articles/10.5802/roia.5>

Céline Leclaire, Lucie Termignon, « Pour une éthique de la recommandation personnalisée à la Bibliothèque nationale de France ». Présenté lors du satellite IFLA *New Horizons in Artificial Intelligence in Libraries*, 21-22 juillet 2022, Galway, Irlande. À paraître.

Philippe Vallas, « Prédire l'état matériel des documents : Dalgocol, un programme de recherche en intelligence artificielle à la BnF : entretien avec Philippe Vallas », *Bulletin des bibliothèques de France (BBF)*, 2022-1. <https://bbf.enssib.fr/consulter/bbf-2022-00-0000-008>