

L'intelligence artificielle et le monde du livre

Livre blanc



TOM GAULD

Tom Lebrun
René Audet

Les recherches et la rédaction de ce livre blanc s'inscrivent dans le cadre du projet de recherche en partenariat Littérature québécoise mobile, financé par le Conseil de recherches en sciences humaines du Canada. Sa réalisation a été rendue possible grâce au soutien de l'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique (OBVIA), financé par le Fonds de recherche du Québec – société et culture.

Ce document est mis à disposition sous licence CC BY-NC 4.0.



Permalien de ce document : <https://doi.org/10.5281/zenodo.4036246>

À propos des auteurs :

Tom Lebrun est juriste, spécialisé en droit du numérique et en droit d'auteur. Étudiant au doctorat en littérature à l'Université Laval, il travaille actuellement sur la question de la génération de textes par intelligence artificielle. Il publie régulièrement sur les questions de droit d'auteur, de culture numérique et de rapport entre droit et intelligence artificielle. Ses recherches sont financées par le Fonds de recherche du Québec – société et culture (FRQSC).

René Audet est professeur titulaire au Département de littérature, théâtre et cinéma de l'Université Laval. Spécialiste de littérature contemporaine et de culture numérique, il s'intéresse aux formes narratives actuelles, au livre et à la littérature numériques, à l'innovation éditoriale et aux enjeux de la diffusion savante en contexte numérique. Il dirige le pôle Québec du projet de recherche en partenariat Littérature québécoise mobile (CRSH), de même que l'éditeur numérique Codicille.

Traduction vers l'anglais : Tiffany Templeton

Graphisme : Hugues Skene, KX3 Communication

Laboratoire Ex situ, Université Laval, septembre 2020.

Dépôt légal, 3^e trimestre 2020.

Bibliothèque et Archives nationales du Québec

ISBN : 978-2-924446-26-3

Image de couverture : © Tom Gauld, « Novel-Writing Algorithm », 2019. Traduction vers le français par DeepL, affinée par un humain.

CRSH  SSHRC



Fonds de recherche
Société et culture
Québec 

Table des matières

Introduction	5
L'intelligence artificielle, qu'est-ce que c'est ?	6
La collecte de données, un passage obligé pour l'utilisation d'IA	6
L'IA dans le milieu du livre: un allié naturel	6
Quelques notions	7
Prédire les attentes des lecteurs	8
Les besoins	8
Les limites	9
L'expertise de Octavio Kulesz	9
Prospectives	9
Synthèse	10
Recommandations	10
Ouvrir l'horizon de la création	11
Prospectives	12
Les besoins	12
Les limites	12
Le cas précis de Fable	12
L'expertise d'Allison Parrish	13
Synthèse	13
Recommandations	14
Assister l'édition	15
Prospectives	15
Les besoins	15
Les limites	16
Le cas précis de QualiFiction et Booxby	16
L'expertise de Virginie Clayssen	17
Synthèse	17
Recommandations	17

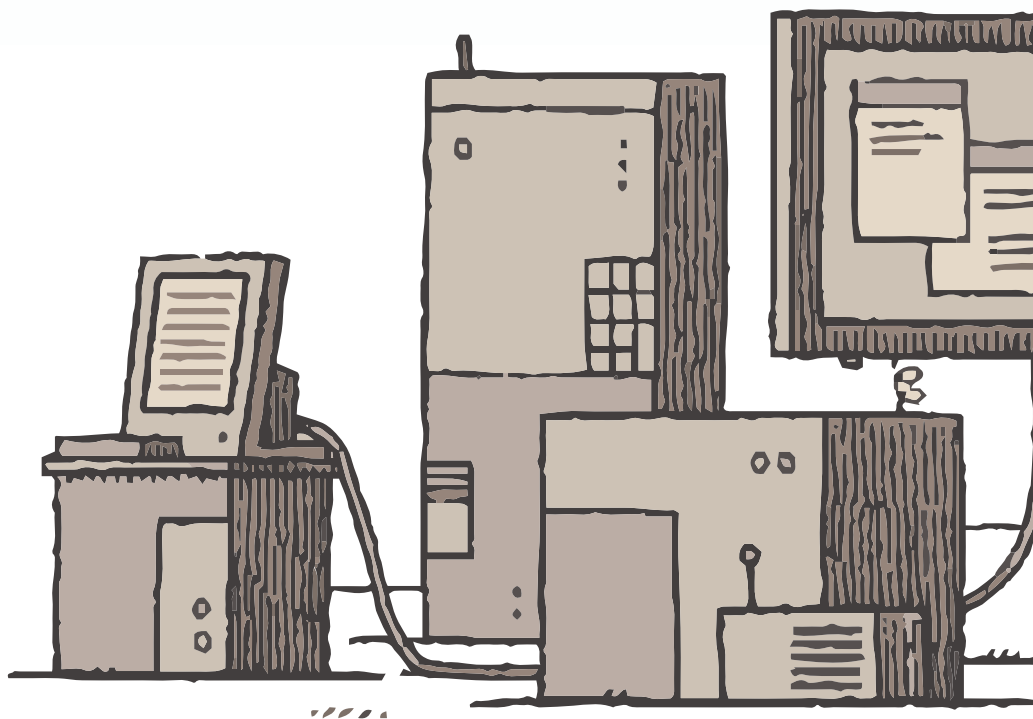
Diffuser et distribuer de façon optimisée	18
Prospectives	19
Les besoins	19
Les limites	19
L'expertise de Véronique Guèvremont	20
Le cas précis de Tamis et StoryFit	20
Synthèse	21
Recommandations	21
Bonifier l'accès au livre	22
Prospectives	22
Les besoins	23
Les limites	23
Le cas de BookGenie, de HarperCollins	24
L'expertise de Clément Laberge	24
Synthèse	25
Recommandations	25
Prospectives littéraires, par David « Jhave » Johnston	26
Conclusion	27
Bibliographie	29

Introduction

À l'heure actuelle, l'intelligence artificielle (IA) fait l'objet de fantasmes et de craintes souvent irraisonnées. Elle trouve à s'insérer dans toutes les sphères de notre société : consommation, médecine, industrie, vie quotidienne... Pourtant, encore peu de travaux ont été consacrés à sa mobilisation par les différents acteurs de l'écosystème du livre. Ce manque est d'autant plus critique que les grands acteurs du numérique – Amazon en tête – s'engouffrent très largement dans cette voie depuis un certain nombre d'années. L'objectif de ce livre blanc est donc d'illustrer les possibilités que les technologies associées à l'IA ouvrent au monde du livre pour ses divers acteurs. Des gestes, des pistes d'action peuvent être mis en œuvre dès maintenant pour tirer parti de cet outil et réagir à une concurrence qui, elle aussi, s'organise.

Après une rapide explication de ce que signifie cette expression « intelligence artificielle », la chaîne du livre sera donc examinée suivant son ordre le plus logique, depuis l'idéation du livre jusqu'à sa vente, en passant par les questions d'écriture, d'édition, de diffusion et de distribution. À mi-chemin entre l'étude positive (au sens de ce qui existe déjà) et prospective (au sens de ce qui sera peut-être), nous avons tenté de fournir autant d'exemples que possible de l'usage des technologies existantes. Les objectifs de ce livre sont donc :

- d'élaborer un aperçu des pratiques d'IA déjà en place dans le monde du livre ;
- de fournir des clés pour comprendre les besoins et les limites de cette technologie par secteur ;
- de suggérer des pistes pour identifier des possibilités ou des perspectives d'affaires fondées sur l'IA.



L'intelligence artificielle, qu'est-ce que c'est?

L'intelligence artificielle (IA) est un terme **générique et englobant**, qui peut recouvrir plusieurs définitions. Les experts s'accordent usuellement à en reconnaître deux principales : la première présente l'IA comme un système capable d'imiter une activité normalement réservée à l'être humain. L'autre définit quant à elle l'IA comme la capacité d'un système à réagir de manière appropriée dans son environnement, autrement dit comme un système capable de fournir la bonne réponse à un problème donné.

Si aucune de ces définitions n'est pleinement satisfaisante, c'est parce que l'IA regroupe en réalité un ensemble de technologies extrêmement diversifiées, qui va du simple correcteur orthographique de traitement de texte des années 1990 à la voiture autonome qui nous conduira peut-être demain. Depuis quelques années, le discours autour de l'IA tend à impliquer surtout ce que l'on appelle l'**apprentissage machine** (*machine learning* en anglais, également appelé apprentissage automatique). Ici encore, il s'agit d'un terme générique qui regroupe un ensemble de technologies diverses, que l'on désigne couramment par le terme d'**algorithmes**. Un algorithme est un peu comme une recette de cuisine : une suite d'instructions à suivre en vue de l'obtention d'un résultat. La particularité de l'IA contemporaine tient au fait que certains algorithmes sont aujourd'hui capables d'apprendre par eux-mêmes de nouvelles tâches. Pour qu'ils y parviennent, il faut les *nourrir* à l'aide d'une grande quantité de **données**.

Si les définitions de l'IA peuvent donc varier, son objectif demeure toujours le même : utiliser des données pour **optimiser** certains processus et/ou en **automatiser** d'autres. En ce sens, l'IA peut servir à découvrir des possibilités autrement invisibles, qui peuvent participer à l'amélioration de l'intelligence d'affaires. Au milieu du livre, l'IA peut donc contribuer d'un grand nombre de manières, notamment par le biais d'applications ou de logiciels spécialisés : pour opérer des traductions, pour transformer des textes en version audio ou encore pour aider à l'édition de manuscrits. Mais dans tous les cas, l'IA nécessite toujours une collecte préalable de données. De beaucoup de données.

La collecte de données, un passage obligé pour l'utilisation d'IA

De façon générale, l'adage veut que plus la quantité de données disponible est élevée, plus l'utilisation d'IA révélera des résultats pertinents. En la matière, le monde du livre est bien mieux servi que la plupart des autres secteurs culturels, à tel point que dans des régions du monde comme le Québec ou la France, la quantité de données disponibles dans certains secteurs de la chaîne du livre pourrait faire croire qu'il est inutile de chercher à en collecter davantage. Cette impression est trompeuse : c'est en **maximisant** le volume de données collectées que nous pourrions **révéler** des perspectives d'affaires qui pour l'instant restent inexploitées. Mais au fait, pourquoi chercher à appliquer des solutions d'IA dans le milieu du livre ?

L'IA dans le milieu du livre : un allié naturel

Tous les milieux économiques et culturels ne sont pas comparables. À cet égard, l'écosystème du livre tient certainement une **place à part**. Souvent héritier de politiques culturelles spécifiques, le monde du livre se distingue en effet par une certaine habitude de la rétroaction entre ses différents acteurs (le libraire informant le diffuseur, ce dernier informant l'éditeur, etc.), voire par une solidarité entre ceux qui œuvrent dans ce créneau culturel au capital symbolique fort. Cette habitude de la **collaboration** s'est souvent organisée grâce à des cadres réglementaires spécifiques, dans un contexte de pressurisation économique croissante provoquée par la mainmise de quelques acteurs de la vente en ligne, particulièrement Amazon. Cette pression est d'autant plus forte que le milieu du livre est très segmenté entre différents types d'édition (savante, littéraire...) et entre de multiples modèles économiques (autoédition, financement

public, entreprises privées), rendant parfois difficile la **planification** de sa réponse aux changements des modèles d'affaires.

Pourtant, l'écosystème du livre résiste et se maintient, sans doute aidé par une conscience partagée de l'idée de culture, où la concurrence est contrebalancée par un certain esprit de communauté. Les réalités vécues quotidiennement sont les mêmes : un cruel manque de temps pour opérer des développements ; la quête constante de ressources humaines, technologiques et financières pour assurer le fonctionnement et encore plus pour envisager des transformations structurelles. Dans un tel milieu, où doivent être menées les réflexions utiles à la planification de l'avenir immédiat de ce créneau culturel et être initiées des actions précises pour asseoir de nouvelles méthodes, de nouveaux modèles, l'idée d'une **concertation** de certains de ses acteurs sur l'utilisation d'IA (voire l'éventuelle mise en commun des données collectées) est donc une possible piste à suivre, que ce livre blanc contribue à esquisser.

Quelques notions

Intelligence artificielle

L'IA est une constellation de technologies diverses, qui tentent historiquement de simuler l'intelligence humaine grâce à l'informatique. Aujourd'hui, parler d'IA revient souvent à parler d'apprentissage machine (*machine learning* en anglais), un champ d'étude fondé sur la statistique et dont l'objectif est de permettre à des systèmes informatiques d'apprendre seuls (autrement dit, sans être explicitement programmés à cet effet) ou de façon supervisée à partir des données qu'on leur fournit. L'apprentissage machine repose pour cela sur différents types d'algorithmes.

Algorithme

Un algorithme est une suite d'opérations conduisant à la résolution d'un problème ou d'une tâche spécifique. Toutes les IA ne font pas usage des mêmes algorithmes : les algorithmes utiles pour la conduite d'une voiture autonome ne sont pas les mêmes que ceux qui sont utilisés en médecine pour la détection de cancers. Dans le monde du livre, l'essentiel des algorithmes utiles relève de ce que l'on appelle le Traitement automatique du langage naturel (TALN) ou *Natural Language Processing* (NLP) en anglais. D'autres types d'algorithmes peuvent toutefois être utilisés, notamment en matière de reconnaissance d'images (pour l'analyse des couvertures, par exemple).

Données

Les données sont l'unité de base de l'information sous forme numérique. Ainsi, une photographie prise par un appareil numérique est un ensemble de données au même titre qu'un livre sous format numérique, tout comme le sont les statistiques de livres empruntés numériquement dans une bibliothèque et l'inventaire d'un entrepôt de livres. La quantité de données qui circulent dans le monde, immense et augmentant de façon quotidienne, nous amène à parler aujourd'hui de l'ère des données massives (*Big data*). Dans le monde du livre, de nombreux types de données peuvent être utiles à l'IA, comme les données statistiques liées à la lecture, les données de vente de livres en librairies ou encore les données relatives au contenu du livre lui-même (s'il contient des scènes crues, si les personnages féminins sont forts, si le langage est soutenu ou familier, si l'action se déroule à Bombay, etc.).

Métadonnées

Les métadonnées sont tout simplement des données qui décrivent d'autres données. Pour prendre l'exemple d'une conversation en ligne, les données correspondent au contenu de la conversation (le contenu du courriel), tandis que les métadonnées sont les informations qui servent à décrire le cadre de cette conversation : les noms et adresses des participants, la taille du message, la date et l'heure de l'échange, etc. Les métadonnées sont utiles pour organiser les connaissances à des fins d'exploitation des données.

Prédire les attentes des lecteurs

Que l'on soit éditeur ou auteur, connaître son public est essentiel pour préparer un livre. L'histoire littéraire est d'ailleurs jonchée de cas célèbres où un éditeur refusait un ouvrage pourtant destiné à marquer son temps, souvent pour des motifs ayant trait au lectorat identifié en amont. Gide a ainsi conseillé à Gallimard de ne pas publier Proust... avant de s'en repentir, quelques années plus tard. En effet, les auteurs et les éditeurs les mieux intentionnés doivent toujours composer avec une **inconnue** majeure quant aux livres qu'ils décident d'écrire et de publier : le texte trouvera-t-il son public ?

On l'a vu en introduction, l'IA contemporaine repose essentiellement sur l'analyse des données qu'on lui fournit. Sans les ériger en exemples, des compagnies comme Netflix utilisent déjà de l'IA sur les données qu'ils récoltent (temps d'écoute, contenus préférés, etc.) de façon à **identifier** ce que le public souhaite voir, avec le succès que l'on sait. La série *House of Cards* est ainsi née du constat qu'il existait un public friand à la fois de Kevin Spacey et de thrillers politiques. Un même système est envisageable dans le milieu du livre. Il est par exemple possible d'extraire les données (achats, listes de souhaits) sur des sites tels que Goodreads afin d'établir une cartographie de ce que le public souhaite lire.

En somme, l'IA peut aider à identifier des tendances de lecture de façon beaucoup plus fine que ce que font les outils habituels d'analyse d'affaires, lesquels se résument souvent à de simples statistiques de vente. En ce sens, l'IA est un outil formidable pour découvrir des **perspectives d'affaires**. Des marchés de niche, qu'un auteur ou un éditeur n'aurait pas identifiés comme intéressants, pourraient ainsi se révéler particulièrement pertinents et trouver leur public (une romance teintée de *fantasy* sur une pop star, par exemple, ou encore un roman féministe qui mettrait en scène Marguerite Duras à l'ère de #metoo).

Les besoins

Bien sûr, rien de tout cela ne peut se faire sans la collecte préalable de **données**, tant sur les œuvres littéraires que sur les lecteurs et leurs pratiques de lecture. Certaines sont déjà disponibles, particulièrement celles des bibliothèques, mais un important travail de concertation entre les acteurs (libraires, éditeurs, auteurs et lecteurs) doit être envisagé si l'on souhaite pouvoir tirer parti au maximum des technologies actuelles d'IA. Ces données doivent être largement **diversifiées** – sur les livres publiés, sur les livres vendus et les livres lus (ce n'est pas la même chose), sur les thématiques, les lieux où se déroulent les intrigues, les époques, voire des données encore plus fines comme le registre de langue, des schémas d'action, le nombre de personnages et leur caractère, etc. Toute information est bonne à exploiter, car c'est en partant du spectre le plus large possible que l'IA permettra de découvrir des champs à explorer et à investir. La condition du succès, à cet égard, est certainement la **cohésion** de la chaîne du livre dans son ensemble. On parle ainsi des données du monde du livre, mais il en existe d'autres à l'ère des **données massives** (*Big data*), notamment celles rendues disponibles par les institutions gouvernementales (les données sur les formations post-secondaires idéales souhaitées par les étudiants en France par exemple, disponible sur data.gouv.fr). Le monde du livre peut s'en saisir afin de les croiser avec ses propres données, ce qui pourrait lui fournir des informations stratégiques, par exemple sur l'orientation culturelle du lectorat jeunes adultes.

Quelques chiffres

Selon une étude de KPMG, 94 % des éditeurs ont conscience de l'importance des métadonnées pour faire évoluer la promotion de leurs catalogues. 77 % des éditeurs ne savent cependant pas de quelle façon faire évoluer leurs pratiques.

Les limites

Le monde du livre est un écosystème particulier. De façon générale, il faut toujours avoir en tête que les productions artistiques qui sortent de la moyenne, les œuvres d'avant-garde seront nécessairement dans l'**angle mort** de l'IA. L'un des risques est ainsi une certaine uniformisation de la production culturelle, car une IA ne réagit qu'en fonction des données qu'elle traite et ne peut extrapoler par rapport à ces dernières. On l'a vu en introduction : **l'IA est un outil qui se sert des données du passé pour tenter de prédire le futur avec des probabilités**. Une IA ne pourrait donc pas recommander à la publication un texte de Proust, car le style d'écriture du célèbre auteur diffère trop de la moyenne stylistique des textes.

L'utilisation perspicace de l'IA consisterait donc à travailler de concert avec de tels systèmes, tout en maintenant un esprit critique : l'IA doit demeurer un **outil d'aide** et non un moyen de remplacer à peu de frais les travailleurs de l'écosystème du livre. En ce sens, l'IA participe davantage de l'**intelligence augmentée** que d'une véritable *intelligence artificielle*.

Prospectives

- L'IA sert à faire des **corrélations** et peut donc servir à identifier des rapprochements pertinents : qu'une majorité des lecteurs d'un certain auteur de romans policiers, par exemple, sont également passionnés de biographies historiques, cela pourrait inspirer la rédaction d'un polar ayant comme héros ou héroïne un personnage historique.
- Une analyse par IA des envies des lecteurs pourrait détecter des intérêts naissants pour une thématique (la fantasy écologique, par exemple) suffisamment en amont pour permettre aux auteurs ou aux éditeurs de créer des œuvres sur le sujet.
- La récolte de données sur le public peut également servir à l'optimisation par IA de la communication et du marketing (en évaluant le potentiel du titre de l'ouvrage sur les ventes auprès de certaines clientèles cibles, par exemple).

L'expertise de Octavio Kulesz

Le potentiel de l'intelligence artificielle (IA) dans le domaine de l'art et de la culture est, sans aucun doute, extraordinaire. Il existe déjà une grande variété de projets d'applications qui proposent des systèmes automatisés conçus pour la co-écriture, la rédaction d'articles, la correction, la traduction, la génération de résumés, l'évaluation de manuscrits et la recommandation de textes. En tout état de cause, les avancées des systèmes IA dans le domaine de l'écriture et de l'édition semblent inexorables, ce qui impliquera à la fois des opportunités et des risques. L'utilisation de ces puissants outils pourrait entraîner une augmentation de la productivité des écrivains et des éditeurs, ainsi que



la généralisation de méthodes marketing plus efficaces pour atteindre le public. Cependant, à moyen terme, la formidable capacité des grandes plateformes numériques à intégrer tous les maillons de la chaîne pourrait transformer ces dernières en acteurs de facto incontournables pour toute activité créative. L'amélioration progressive des outils IA entre les mains d'un petit nombre d'entreprises pourrait ainsi avoir un impact négatif sur l'emploi, remplaçant les correcteurs et les traducteurs mais aussi les journalistes, les éditeurs et les écrivains eux-mêmes. Dans les années à venir, pour affirmer un paysage littéraire et éditorial au sein duquel pluralisme et diversité auraient augmenté, il est essentiel d'une part de protéger les écosystèmes culturels dont la disparition pourrait entraîner une perte irréversible, et d'autre part de favoriser l'émergence de nouvelles applications et startups IA locales, évitant ainsi la concentration de l'offre.

Octavio Kulesz est le directeur de la maison d'édition Editorial Teseo, à Buenos Aires. Il travaille comme chercheur dans les domaines relatifs aux médias sociaux et à la culture numérique dans les économies émergentes et agit comme expert pour l'UNESCO sur la Convention sur la protection et la promotion de la diversité des expressions culturelles.

Synthèse

Analyser et cartographier des envies de lecture **qui n'existent pas encore** est un des grands possibles de l'IA dans le monde du livre. On peut déjà envisager la création d'outils ciblant des thèmes de lecture d'une manière inédite, non seulement auprès du grand public, mais aussi pour des publics de niche, en recherche de textes plus spécialisés. L'IA pourrait ainsi aider à identifier l'intérêt d'écrire et de publier certains ouvrages singuliers, alors qu'autrement ils seraient intuitivement laissés de côté par les professionnels du livre. Pour y parvenir, la récolte de données doit impérativement être **enrichie** et ne pas se limiter aux seules métadonnées (le genre du livre, le nom de l'auteur, sa date de publication, etc.). Les grands acteurs de la culture en contexte numérique (comme Netflix et Amazon) vont sans conteste – si ce n'est pas déjà le cas – se saisir de ces possibilités offertes par l'IA en amont des créations artistiques. Il est donc fondamental que l'écosystème du livre s'y investisse **le plus tôt possible**.

Quelques chiffres

Selon un article du *New Yorker*, l'algorithme présenté dans l'ouvrage *The Bestseller Code* permet, après une analyse des textes, de prédire avec un taux de succès de 80% leur appartenance à la liste des meilleures ventes du *New York Times*.

Recommandations

- Les acteurs du secteur du livre possèdent déjà des lots significatifs de données, ce qui constitue un actif important à investir. Il est primordial de **diversifier et augmenter** significativement la collecte de données : la plus grande diversité est requise, que ce soit sur les thèmes de lecture du public ou les ouvrages actuellement en circulation (thématiques, lieux, époques à laquelle se déroule l'intrigue, etc.). Cette diversification des données conduira à cartographier plus finement les envies des lecteurs, ce qui servira ultimement à toute la chaîne du livre.
- Les possibilités de collecte, de conservation et de stockage doivent **être réfléchies en amont de la chaîne du livre** pour ne pas être capturées par des acteurs trop importants (lesquels ne redistribueraient ni les données, ni les compétences acquises et encore moins les résultats d'analyse). Les acteurs institutionnels, les bibliothèques et les associations ont leur rôle à jouer en la matière.
- Il est primordial de toujours évaluer les solutions proposées par l'IA avec un esprit critique. L'IA est un outil conservateur par nature, puisqu'il se fonde toujours sur les données du passé : c'est son usage qui déterminera son impact futur sur le milieu du livre.



Ouvrir l'horizon de la création

Une IA peut-elle écrire ? La question n'est pas aussi saugrenue qu'elle peut paraître de prime abord : elle l'est même de moins en moins, comme on l'observe déjà au quotidien dans le monde du journalisme et des communications. Au-delà du fantasme, beaucoup d'outils actuels permettent déjà d'écrire. Le correcteur orthographique de votre traitement de texte ou de votre téléphone intelligent relève déjà de ce qu'on appelle l'intelligence artificielle, de même que les outils de prédiction de mots sur certaines de vos messageries courriel. Ces outils ne sont cependant que des aides à l'écriture. D'autres, beaucoup plus performants (et gourmands en énergie et en données) **écrivent littéralement des textes...** Et parfois des fictions et de la poésie, avec des succès variables selon les conditions et nos attentes.

La compagnie OpenAI a développé en 2019 une IA nommée GPT-2, laquelle génère des textes de fiction ou de journalisme à peu près crédibles pour une lecture rapide et superficielle. Le système GPT-2 a causé **quelques remous** dans le milieu de l'IA, car il était assez généralement admis que la maîtrise du langage naturel (et donc de l'écriture) était encore un objectif lointain. Si les textes générés par les algorithmes en question ne sont pas parfaits – une lecture fine des exemples fournis par OpenAI met en évidence des incohérences logiques, entre autres –, leur progrès spectaculaire **impressionne tout autant qu'il suscite des interrogations**. De nombreux outils gratuits et faciles d'utilisation permettent du reste de s'en faire une idée, sans qu'aucune connaissance informatique soit nécessaire :

- Scratch Pad d'Open AI
- AI Dungeon
- Write With Transformer (WWT)

Très récemment, la compagnie a sorti une nouvelle version de son système, GPT-3, qui est encore plus performante. Au vu des tests actuellement effectués (leur interface de programmation a été rendue disponible au public, sur liste d'attente), la cohérence des textes générés semble accrue. Difficile de prévoir les progrès futurs en matière de génération textuelle, mais s'il est peu probable que de tels textes soient demain reçus comme des chefs-d'œuvre de littérature, ils pourraient toutefois s'insérer avantageusement dans un **marché de masse** où la récurrence de modèles narratifs est de mise. Les romans Harlequin, pour citer un exemple bien connu, ont largement illustré la rentabilité de ce créneau, bien avant la venue de l'informatique.

En marge de tels usages, il est plus probable que l'IA soit utilisée comme **assistant d'écriture**. Certains auteurs en font déjà une utilisation littéraire : au Japon, un manga généré par IA intitulé *Paigon* est le fruit de la capture esthétique des travaux de Osamu Tezuka, le père du fameux Astro Boy. En Californie, le logiciel Granthika permet à son créateur, l'écrivain primé Vikram Chandra, de cartographier les éléments narratifs du texte pendant sa création, pour mieux se retrouver dans les schémas narratifs complexes et dans la chronologie des personnages. De son côté, le poète montréalais David «Jhave» Johnston utilise l'IA pour générer des premiers jets qu'il retravaille ensuite en tant que poète. Autre cas de figure, le romancier Robin Sloan use de son côté l'IA pour générer les dialogues de l'un de ses personnages (une IA, en l'occurrence !).

Quelques chiffres

Le système de génération actuellement le plus performant requiert 40 Go de texte pour fonctionner. À titre indicatif, cela représente environ 13 000 fois la totalité d'*À la recherche du temps perdu* de Marcel Proust.

Écrire un texte ne représente donc que la partie la plus directement visible et spectaculaire du **spectre des possibilités offertes par l'IA** dans le milieu du livre. L'IA peut également être utilisée pour rédiger des résumés sur mesure de romans en fonction des lecteurs potentiels, de leurs goûts et de leur sensibilité, ou pour orienter le design des livres ou même leur illustration. *Machines Upon Every Flower* (Gregory Chatonsky et Karmel Allison, Anteism Press, 2018) recueille par exemple des poèmes générés par IA à partir d'un mot, lesquels poèmes seront ensuite interprétés par une seconde IA pour produire des images. Autre exemple d'utilisation, *1 the Road* de Ross Goodwin se présente comme un « roman » écrit par une voiture lors d'un trajet de Brooklyn à la Nouvelle-Orléans – une version intégrale (non corrigée/éditée) a été publiée par Jean Boîte éditions en 2018... Encore une fois, tout reste à inventer.

Prospectives

- L'IA pourrait conduire à développer des assistants créatifs personnifiés, de façon à aider à écrire ou même de dessiner « à la manière » d'un auteur, par exemple.
- L'IA peut permettre de découvrir ou de créer de nouvelles possibilités de création, notamment au travers du croisement de données (par exemple, une application propulsée par IA pourrait rédiger automatiquement des carnets de voyage en commentant les photographies de famille).
- Croiser les données du livre, les données ouvertes et celles de l'internet des objets (*Internet of Things*) de demain pourront également dessiner de nouvelles avenues. La lecture d'un livre de cuisine pourrait suggérer des recettes sur l'écran d'un réfrigérateur connecté et passer automatiquement la commande auprès d'un supermarché.

Les besoins

Comme pour la cartographie des envies de lecture, l'écriture par IA nécessite des données, beaucoup de données. Des quantités impressionnantes de texte doivent être ainsi analysées par les systèmes d'IA, ce qui entraîne des **coûts importants** en termes de puissance de calcul, en préparation et en opération par du personnel qualifié, mais aussi pour la validation et le calibrage des données. L'avantage de ces systèmes est cependant qu'ils sont **reproductibles** : s'ils sont mutualisés, leurs coûts de développement peuvent être considérablement amortis.

Quelques chiffres

L'écriture générée par IA est également utilisée dans le journalisme. Environ un tiers des nouvelles publiées par Bloomberg est ainsi générée par des machines.

Les limites

Afin de prévenir toute infraction et en l'absence de loi ou de jurisprudence stable sur la question de l'utilisation d'IA sur des données protégées par le droit d'auteur, il est tout indiqué pour les éditeurs et les auteurs de privilégier le recours à des **contrats types**, lesquels encadrent explicitement le minage de ces œuvres à des fins d'analyse pour développer et améliorer des solutions propulsées par IA. Sans de tels contrats, il est recommandé de recourir à des données relevant soit du domaine public, soit celles qui sont régies par des licences Creative Commons permettant leur réutilisation et leur modification.

Le cas précis de Fable

Extraire vos personnages de romans préférés pour les faire vivre numériquement et surtout pour pouvoir communiquer avec eux : c'est le pari de la compagnie Fable. En multipliant les supports possibles de communication (en réalité virtuelle, mais aussi à terme sur téléphone intelligent, ordinateur ou tablette), l'entreprise souhaite renverser le caractère trop souvent impersonnel des assistants virtuels que nous utilisons déjà (Siri, Alexa, Google Assistant, etc.). Cette stratégie pourrait conduire à un **engagement émotionnel** accru avec des personnages ayant leurs qualités comme leurs défauts. Dans votre salon, vous pourrez ainsi converser avec le capitaine Achab de *Moby Dick*, demander l'avis d'Hermione Granger sur la politique américaine ou encore (ça fait frémir) laisser les monstres de Stephen King répondre à votre messagerie vocale au moment d'Halloween.

L'expertise d'Allison Parrish

En 2020, l'utilisation d'intelligence artificielle dans l'écriture littéraire est tout à fait courante. Des technologies telles que le correcteur orthographique, l'autocomplétion et les claviers prédictifs (comme le QuickType d'iOS), chacune d'elles relevant de l'«intelligence artificielle», sont omniprésentes dans les logiciels de traitement de texte autant que dans les applications de messagerie instantanée et de courriel. Presque tous les textes publiés aujourd'hui sont composés à l'aide d'outils de ce type. Il n'est donc pas déraisonnable de prétendre que presque tous les textes publiés sont «co-rédigés» par des intelligences artificielles. Pour cette raison, l'idée même de «texte généré par ordinateur» est en soi incohérente et témoigne d'une incapacité à identifier le contexte de rédaction de tout texte, et de là le travail humain qui lui est associé. Malgré tout, la question de l'attribution de l'autorité plane quand on s'interroge sur les usages de l'intelligence artificielle dans l'écriture. En particulier, les grands modèles de langage basés sur des réseaux neuronaux comme OpenAI GPT-3 (ou son prédécesseur, GPT-2) – qui semblent produire de manière autonome de longues portions de texte cohérent – suscitent de profonds questionnements philosophiques sur l'identité de l'auteur, alors que les technologies quotidiennes comme le correcteur orthographique et les claviers prédictifs ne les suscitent aucunement. Pourtant, il y a peu de différence entre les deux du point de vue de l'autorité. Le texte produit par GPT-3 ne peut ainsi être compris comme «généré par la machine» que dans le sens banal où un courriel est «co-rédigé» par la fonction de vérification orthographique d'un logiciel de messagerie. Autant pour la vérification orthographique que pour GPT-3, il importe de reconnaître que ces systèmes sont créés par des personnes, résultent de décisions prises par des humains et sont porteurs de biais humains. Par exemple, les ingénieurs de GPT-3 ont fait des choix quant à son architecture, ils ont sélectionné les textes à inclure pour l'entraînement du modèle et ils ont choisi de quelle manière ils allaient mettre leurs outils à la disposition des autres. Je soutiens que les textes produits par GPT-3 ne sont pas écrits par une «machine» (comme on le prétend souvent), mais sont plutôt coécrits par les ingénieurs du modèle, par les auteurs dont les ouvrages ont été mobilisés dans le processus d'entraînement et par l'utilisateur final qui utilise le texte généré à des fins particulières. En tant que poète, mon but est d'inventer de nouvelles formes de langage. Je suis intéressée par l'utilisation de l'intelligence artificielle non pas pour lui transférer l'autorité d'un texte, mais bien pour diffracter cette autorité et la rendre problématique. Les poètes ont utilisé des techniques statistiques et informatiques à cette fin depuis au moins un siècle, depuis *Pour faire un poème dadaïste* de Tristan Tzara en 1920, jusqu'au remarquable *Travesty Generator* de Lillian-Yvonne Bertram publié en 2019. J'ai la conviction que les techniques actuelles d'intelligence artificielle et d'apprentissage machine ont un grand potentiel pour perpétuer cette tradition.



Allison Parrish est une programmeuse, poète, éducatrice et conceptrice de jeux vidéo dont l'enseignement et la pratique s'intéressent aux phénomènes inusuels qui surviennent lorsque le langage et l'informatique se rencontrent. Elle est l'autrice de différentes œuvres générées par informatique, notamment *Articulations* chez Counterpath Press (2018).

Synthèse

Pour l'heure, en contexte littéraire, l'IA est mobilisée davantage comme un outil d'écriture pour compléter ou corriger des textes que comme agent d'écriture autonome. Il est peu probable que cette situation évolue radicalement dans les années à venir, même si ce n'est jamais à exclure. Ce point est rassurant, au sens où l'IA continue encore de relever, au moins pour ce qui concerne l'écriture, davantage de l'intelligence augmentée que d'une véritable conscience artificielle. Il n'empêche que des applications tirant parti de l'IA peuvent d'ores et déjà permettre d'investir de façon **novatrice** les champs de l'écriture, mettant au jour des dynamiques d'écriture inédites ou offrant des pistes pour le futur du livre. C'est déjà le cas en matière de journalisme, où l'IA produit de nombreux articles qui résument couramment des résultats sportifs, financiers ou politiques.

Recommandations

- Des contrats types autorisant l'utilisation des œuvres protégées par le droit d'auteur dans le cas de la recherche en IA doivent être mis en place. La plus-value de telles autorisations serait par ailleurs d'une importance capitale pour établir une stratégie cohérente sur l'IA dans le monde du livre.
- Le développement d'outils de création ou d'animation par IA nécessite beaucoup de données et de puissance de calcul. La chaîne du livre a ainsi tout intérêt à **mutualiser ces coûts**, qui peuvent aussi être financés en tout ou en partie par les institutions publiques afin de bénéficier aux plus petits acteurs.



Assister l'édition

S'il est un milieu de l'écosystème du livre qui peut réellement être transformé par l'utilisation d'IA, c'est bien le secteur éditorial. Sans que l'on puisse proprement parler de révolution, l'IA peut en effet participer à l'amélioration et à l'optimisation d'un grand nombre des tâches habituelles des éditrices et des éditeurs. On préférera là aussi parler d'**intelligence augmentée** : le rôle de chaque individu au cœur du processus éditorial doit rester primordial pour l'édition de demain. Il n'en demeure pas moins que la panoplie d'outils que constitue l'IA peut déjà accomplir un certain nombre de fonctions.

En l'état, l'IA peut donc servir à **optimiser considérablement** le travail d'une maison d'édition, libérant du temps et de l'argent pour d'autres tâches. En plus de **mieux référencer** les titres d'une maison (pour bien sûr mieux les distribuer par la suite), l'aide à la traduction pourra ainsi produire de premiers jets dans des langues auparavant inaccessibles, divisant d'autant le temps, et donc le coût, de tels investissements. La lecture des manuscrits par IA peut également permettre de faire un premier classement des textes proposés, par exemple en les triant en fonction de leur genre, facilitant l'attribution des manuscrits aux lecteurs de la maison d'édition. En plus d'identifier le moment où la publication du titre serait la plus efficace, des **prédictions de vente** plus précises peuvent être modélisées, autorisant la prise de risque sur d'autres titres plus exigeants et moins rentables et participant d'une publication du nombre d'exemplaires plus ajustée. D'autres tâches peuvent encore être automatisées, comme l'attribution des redevances (royautés) en fonction des ventes, la détection des infractions au droit d'auteur ou encore le formatage des textes selon le protocole de la maison d'édition.

Quelques chiffres

Les éditeurs reçoivent un très grand nombre de manuscrits, dont seulement 2 % seront publiés en moyenne, parmi lesquels 1 % deviendront des best-sellers. Faire un premier tri dans ces manuscrits grâce à l'IA, ne serait-ce que pour classer les titres en fonction de leur thème, procurerait un gain de temps considérable.

Prospectives

- Certains assistants intelligents recommandent déjà des ouvrages grâce à l'interaction vocale. La compagnie Simon & Schuster a ainsi doté le système Alexa d'Amazon d'une « librairie Stephen King » qui propose des titres de l'auteur en fonction des préférences de l'utilisateur, l'assistant vocal allant jusqu'à poser des questions pour mieux se calibrer. Ce système pourrait être étendu et ainsi recommander des ouvrages d'autres auteurs.
- Les très rapides progrès de l'IA dans les domaines comme l'ingénierie sonore permettent d'imaginer qu'il sera bientôt possible de **transformer des manuscrits en livres audio de façon automatisée**. En Chine, la compagnie Sogou automatise déjà la lecture de textes par IA en reproduisant la voix des auteurs. En Occident, des compagnies comme Lyrebird travaillent dans le même sens. On peut donc envisager sous peu l'existence d'une application qui rend possible de sélectionner notre lecteur préféré, comme un acteur ou même un membre de notre famille, pour nous lire un ouvrage qui vient de paraître.
- Les possibilités offertes par l'IA peuvent d'ores et déjà améliorer les processus d'autoédition. Déjà utilisées par les grandes plateformes pour la mise en forme des manuscrits, des applications traduiraient des textes dans de nouvelles langues, donnant ainsi aux auteurs autoédités l'accès à des marchés auparavant inatteignables.

Les besoins

Les éditeurs occupent une position **privilégiée** pour la récolte de données dans le milieu du livre, tant sur les titres produits que sur les publics auxquels ils sont vendus. L'organisation du milieu, mais aussi et surtout la présence de moyens financiers et humains conséquents les désignent comme candidats naturels au développement de solutions IA. Comme tout système d'IA nécessite une grande quantité de données, on peut supposer que les plus grandes maisons d'édition auront un avantage comparatif indéniable sur les plus petites, eu égard à l'ampleur de la production éditoriale – et donc à la récolte de données – à leur disposition. Cette impression ne doit cependant pas faire oublier qu'un certain nombre d'entreprises offrent déjà des services d'IA à des coûts **accessibles** pour les petites structures. Que ce soit de manière externalisée ou à l'interne, l'IA dans le monde de l'édition est donc une **réalité**.

Quelques chiffres

Selon l'Association of American Publishers, le livre audio est le format qui connaît actuellement la croissance la plus forte (pas moins de +36,5% lors des deux premiers mois de 2019). Un développement des outils de transformation des manuscrits en livres audio serait une façon pertinente de tirer des bénéfices substantiels de ce mouvement de fond.

Les limites

En déterminant le marché potentiel pour un ouvrage donné, l'IA participe indirectement à l'édition du texte. En effet, un outil d'analyse par IA pourrait suggérer que le manuscrit à éditer est en l'état trop long pour son lectorat potentiel, qu'il faut réduire le nombre de pages par chapitre ou encore que les contenus offensants doivent être révisés. De la même manière, l'IA pourrait suggérer de ne pas publier des ouvrages dont la valeur littéraire est pourtant importante : en tant qu'outil basé sur la statistique, elle ne peut évidemment établir de jugement esthétique. Un système automatisé rejeterait ainsi clairement la plupart des grands auteurs des derniers siècles, que ce soit en raison des phrases trop longues, de la syntaxe trop complexe ou encore du manque de résolution de l'intrigue. Il en va de même en matière de traduction : l'IA n'égale sans doute jamais le travail créatif des meilleurs traducteurs et traductrices dans leur domaine, même si elle peut fournir des premiers jets rapidement et à peu de frais. En résumé, le classement par IA, son analyse des marchés tout comme son importance comme outil d'assistance en matière éditoriale sont donc toujours à prendre avec des pincettes : il faut idéalement **maintenir une décision humaine** en amont (pour la configuration ou encore le choix des algorithmes) et en aval des traitements automatisés envisagés.

Quelques chiffres

Le nombre de livres numériques autopubliés représentait 34% du marché américain du livre numérique en 2017. Ce secteur de l'autopublication bénéficie encore peu des technologies d'IA.

Le cas précis de QualiFiction et Booxby

Les compagnies QualiFiction et Booxby proposent d'utiliser l'IA pour assister les éditeurs dans l'évaluation des manuscrits. Selon Elisabeth Mol de la revue en ligne *Lettres numériques*, le logiciel LiSA qu'a développé la compagnie QualiFiction analyse ainsi un manuscrit avant de fournir à l'éditeur un rapport sur un grand nombre de paramètres, tels que le niveau

de suspense ou de réflexion, la complexité des phrases ou encore le niveau d'innovation par rapport au reste du catalogue de l'éditeur. En plus de ses capacités de **reconnaissance émotionnelle** – qui lui permettent d'indiquer les sentiments ressentis par le lecteur au fil du livre ainsi que leur taux de présence –, le système s'adapte à la ligne éditoriale de la maison d'édition, écartant par exemple tous les romans jeunesse si cette dernière ne publie que des polars. LiSA va jusqu'à proposer une courbe qui représente la structure et la dramaturgie du manuscrit, attribuant au texte ni plus ni moins qu'un « best-seller score »... La société Booxby, quant à elle, propose sensiblement la même chose : analyser un texte afin de déceler la meilleure stratégie marketing pour en faire la promotion auprès du public.

Synthèse

L'utilisation d'IA dans le monde de l'édition relève d'une certaine évidence, tant les possibilités d'optimisation de certaines tâches conduisent à dégager du temps et de l'argent, autant **d'atouts précieux** dans un secteur culturel sous pression. En la matière, il est certain qu'Amazon et d'autres acteurs du même type capitalisent déjà sur les possibilités offertes par l'IA. L'automatisation du classement des manuscrits, l'aide à la traduction, la planification fine de la mise en marché ou encore les possibilités en matière de transformation des titres en livres audio apparaissent comme autant de **promesses d'évolution** pour le secteur qui ne doivent pas être laissées aux seuls grands joueurs.

Recommandations

- Les informations acquises sur les catalogues grâce à leur traitement par IA sont des données précieuses. En les partageant avec d'autres acteurs de l'écosystème du livre (par exemple par le biais d'accords sectoriels), ces données pourraient devenir le moteur de considérables **retours sur investissement** pour l'édition. En la matière, la mutualisation des données est la promesse d'une synergie économique à envisager sérieusement.
- Pour promouvoir la collecte des données et sa diffusion éventuelle auprès des autres acteurs de l'écosystème du livre, un investissement du **secteur public** en la matière serait sans aucun doute pertinent. Inciter les acteurs à partager leurs données, que ce soit par la mise à disposition d'outils *open source* ou par le biais d'incitatifs financiers (crédits d'impôts, subventions à l'innovation, etc.) est également une voie qui gagnerait à être explorée.



L'expertise de Virginie Clayssen

Je peux témoigner d'usages que j'ai eu l'opportunité d'observer directement ou de voir présentés. L'un concerne la découvrabilité. Des technologies basées sur l'IA peuvent en effet contribuer à la classification a posteriori d'ouvrages (dans le cas d'une numérisation massive) en générant des métadonnées pertinentes. Je pense à la mission effectuée par la société Pythagoria pour la société FeniXX, qui gère pour le compte des éditeurs les livres numériques issus de la numérisation des livres indisponibles du XX^e siècle. Cette mission a produit des résultats concrets, qui ont grandement facilité la production de métadonnées pour les livres concernés par la numérisation, ainsi que leur classification, permettant d'offrir à ces livres une meilleure découvrabilité. Une autre tâche que l'on souhaiterait pouvoir confier à des dispositifs basés sur de l'IA a trait à l'accessibilité des livres numériques. L'une des tâches les plus lourdes à effectuer,



pour qu'un livre numérique soit nativement accessible à un public en situation de handicap visuel, est la nécessité de décrire les images lorsque le livre en contient. Autant on peut imaginer instaurer progressivement le réflexe pour les auteurs ou les éditeurs d'inclure en amont du processus de telles descriptions dans les nouvelles publications, autant la tâche consistant à ajouter de telles descriptions à l'ensemble des titres déjà publiés et numérisés nécessite le recours à une automatisation qui se doit d'être la plus complète possible – ce que permet l'IA. Des projets en ce sens ont été menés par la fondation LIA (Libri Italiani Accessibili) en Italie. EDRLab en France souhaite poursuivre les recherches dans cette direction. Si l'IA appelle une vigilance constante concernant son utilisation (par des réflexions critiques, par exemple, sur le « digital labour »), il paraît important pour les acteurs du monde du livre de rester très attentifs à ces avancées technologiques, tant à ce qu'elles pourraient bousculer qu'à ce qu'elles pourraient apporter.

Virginie Clayssen est la directrice du patrimoine et de la numérisation chez Editis, présidente de la Commission Numérique du Syndicat national de l'édition et de l'EDRLab (European Digital Reading Lab). Architecte de formation, elle intervient régulièrement dans des conférences sur l'édition numérique, en France et à l'étranger.

Diffuser et distribuer de façon optimisée

Bien que ces deux secteurs soient usuellement liés, l'impact potentiel de l'intelligence artificielle n'est pas le même pour la diffusion (la dimension commerciale, de représentation et de publicité des livres) que pour la distribution (la dimension logistique, d'entreposage et de livraison de ces derniers).

En matière de **diffusion** – laquelle peut parfois être assumée directement par l'éditeur selon les secteurs –, le potentiel d'utilisation de l'IA a surtout trait à l'automatisation des systèmes de recommandation. À titre d'exemple, le *New York Times* et le *Wall Street Journal* utilisent déjà des systèmes d'apprentissage machine pour adapter leur lettre d'information à chaque lecteur en fonction de ses préférences et de ses usages de lecture. Sans surprise, des procédures similaires peuvent être envisagées en matière de diffusion. Par exemple, la **grille d'office**, qui est le formulaire du diffuseur (d'une maison d'édition, par exemple) déterminant quels livres seront envoyés aux détaillants (librairies, grandes surfaces, etc.) peut être automatisée, au moins en partie, grâce à l'IA.

On peut ainsi sans difficulté imaginer qu'un diffuseur passe un contrat avec différents libraires pour avoir accès à certains sous-ensembles de leurs données pour les exploiter conjointement. Par ce moyen, le diffuseur pourrait alors connaître en temps réel les tendances de lecture dans certaines librairies qu'il dessert et ainsi agir (manuellement ou automatiquement) sur les envois d'ouvrages qui correspondent aux besoins ou aux tendances de lecture chez ces libraires, ce qui aidera ces derniers à mieux gérer leurs stocks. Si les données concernant les livres sont suffisamment affinées et qu'elles rendent possible le classement des ouvrages en fonction de critères extrêmement fins (comme le lieu où se passe l'intrigue ou le fait que le héros ou l'héroïne ait un chien, qu'il n'y ait pas de violence, etc.), alors les outils des diffuseurs seront d'autant plus efficaces et pertinents pour aider les libraires dans leur fonction de conseil. Le diffuseur maintiendra bien sûr son rôle de gestion stratégique de la diffusion, mais bénéficiera éventuellement d'un avantage comparatif indéniable au travers de l'exploitation par IA des données disponibles.

À l'inverse, en matière de **distribution**, l'incursion massive et disruptive d'IA est inéluctable. Alors que les systèmes déjà informatisés se basent souvent sur l'analyse de données par séquence (on opère ainsi un « carottage » des informations relatives au fonctionnement afin d'apporter ensuite les modifications nécessaires), les algorithmes actuels permettent de s'adapter à la situation en temps réel, autrement dit de s'adapter aux flux entrants et sortants d'information. Pour la chaîne de distribution du livre, cela signifie qu'il serait possible d'automatiser en temps réel les commandes de livres depuis les détaillants jusqu'aux distributeurs et aux imprimeurs, voire même jusqu'aux éditeurs. En 2012, la compagnie Amazon a ainsi déposé un brevet concernant les envois anticipés : pour éviter notamment les bris de pièces, qui peuvent entraver le bon fonctionnement de toute la chaîne de l'entreprise, le système anticipe par probabilités qu'une pièce doive être réparée à un moment donné et peut ainsi prévoir de la réparer avant qu'elle brise. De son côté, la compagnie chinoise JD.com utilise des drones non seulement pour livrer dans des endroits très reculés, mais aussi pour automatiser les transferts de marchandises entre ses entrepôts, certains drones transportant jusqu'à cinq tonnes de marchandise.

Quelques chiffres

Selon le rapport 2019 du Syndicat national de l'édition en France (SNE), le tonnage de livres transporté par les distributeurs sur une période de 2015 à 2017 était de 202 900 tonnes pour le flux aller et de 50 370 tonnes pour le flux retour (soit les invendus), un chiffre pouvant être radicalement abaissé grâce à l'optimisation de la distribution et de la diffusion par IA. L'impression locale (par petits tirages) de même que l'impression à la demande sont déjà des solutions potentielles pour accompagner ce mouvement de **minimisation** des flux matériels, par ailleurs avantageux d'un point de vue écologique.

De la réception et de l'entreposage des livres à l'exécution des commandes et leur livraison en passant par la facturation des marchandises, tous les aspects de la distribution peuvent profiter d'une automatisation par l'IA et la robotique. Cette réalité peut constituer une occasion pour le secteur de la distribution : un système d'IA pensé à la fois pour la diffusion et la distribution, raccordé aux données en amont et en aval de la chaîne du livre, pourrait certainement optimiser le fonctionnement du secteur – en réduisant les fameux 25% à 35% de retours usuels, par exemple. Mais pour cela, une concertation des acteurs sur la question de la collecte et de la mise à disposition des données est nécessaire.

Quelques chiffres

Toujours selon ce même rapport, 419 millions de livres ont été vendus en France en 2018. Ces 419 millions représentent 106 799 titres, parmi lesquels 44 968 nouveautés et 61 831 réimpressions. L'IA peut permettre de **naviguer** facilement dans cet ensemble de titres, aidant considérablement le travail de diffusion.

Prospectives

- La grille d'office peut être gérée par IA afin de s'adapter en temps réel aux besoins des détaillants et ainsi baisser ou augmenter la quantité de livres (par exemple en générant des prénotés adaptés aux envies des clients de chacune des librairies d'un secteur géographique).
- Grâce aux données collectées auprès des détaillants, les diffuseurs pourraient cibler bien plus avantageusement le type de publicité à favoriser.
- Pour la diffusion et le marketing, l'IA rend possible de trouver des titres comparables à celui dont on doit assurer la promotion et proposer une stratégie de communication plus adéquate, de sorte de réduire considérablement le taux de retours usuels.

Les besoins

Pour être utile au secteur de la diffusion et de la distribution du livre, l'IA doit ici aussi disposer d'une grande quantité de données. Ce double secteur a ainsi particulièrement intérêt à accéder aux données des détaillants (celles des libraires sur la fréquentation, mais aussi les données sur les lecteurs) et aux données des éditeurs (celles qui détaillent au plus près le contenu même des livres). Des accords doivent donc être réfléchis en amont, ce qui passe par une **concertation** préalable des partenaires éventuellement intéressés, qu'ils soient diffuseurs/distributeurs, détaillants, éditeurs, etc. Il n'est pas obligé que tous les acteurs du milieu décident d'y participer, même si en la matière le nombre fait certainement la force : cette asymétrie potentielle sera évidemment à prendre en compte pour l'établissement de stratégies cohérentes quant à l'introduction des solutions d'IA éventuellement retenues.

L'existence de normes uniformes, comme le format ONIX pour la transmission des données dans la chaîne d'approvisionnement, participera sans aucun doute de cette dynamique.

Quelques chiffres

Selon une étude de la firme McKinsey, les possibilités de personnalisation en matière de communication numérique – notamment par IA – peuvent réduire jusqu'à 50% les coûts d'acquisition, mais aussi augmenter l'efficacité du marketing de 10% à 30%.

Les limites

Dans le milieu du livre, bien distribuer ne veut pas dire distribuer de façon uniforme sur tout le territoire. Si l'IA peut aider à cibler au mieux de quelle manière le livre doit être diffusé puis distribué, il est risqué de la rendre autonome dans tous ses choix. Des auteurs auxquels on croit ou qui ouvrent de nouvelles perspectives en littérature, sans même parler de nouvelles perspectives de marchés, pourraient ne pas être correctement diffusés et distribués par des systèmes automatisés, tout simplement parce qu'ils n'appartiennent pas encore au cadre de référence sur lequel se fonde tout système

L'expertise de Véronique Guèvremont

L'introduction et l'utilisation de l'IA dans la chaîne du livre soulèvent de multiples interrogations sur les répercussions que ce phénomène peut avoir sur la diversité des expressions culturelles et sur la nécessaire adaptation des instruments d'action publique dans ce domaine. D'une part, le recours à l'IA suppose l'accessibilité à une masse de données que tous les acteurs ne sont pas nécessairement disposés à partager. Par exemple, les données massives sur les goûts et habitudes des lecteurs sont recherchées par des auteurs et éditeurs de livres afin de générer



et commercialiser des œuvres susceptibles d'attirer l'attention d'un vaste public. Cependant, certains détenteurs de données – comme les bibliothèques publiques – s'inquiètent légitimement de l'impact d'une mutualisation de telles informations sur la création de nouveaux contenus et l'évolution de la diversité culturelle. D'autre part, l'IA et les algorithmes de recommandation peuvent être de formidables outils pour favoriser la découvrabilité d'une diversité d'œuvres littéraires dans l'environnement numérique. Or,

pour bien exploiter ce potentiel et développer les algorithmes appropriés, il est nécessaire d'avoir accès aux données générées par les lecteurs. Il est aussi important que ces données reflètent les tendances nationales, régionales et locales, afin d'éviter que la découvrabilité ne repose que sur des données générées par de grandes plateformes, ce qui produirait l'effet inverse d'uniformiser les goûts et habitudes de lecture en raison d'une mauvaise utilisation des systèmes de recommandations. À cet égard, une intervention de l'État dans le secteur du livre pourrait être souhaitable en vue de soutenir, mais aussi d'encadrer par le biais d'instruments juridiques appropriés, la collecte, la mutualisation, l'analyse et l'utilisation des données par les différents maillons de la chaîne de valeur du livre.

Véronique Guèvremont est titulaire de la Chaire UNESCO sur la diversité des expressions culturelles et professeure titulaire à la Faculté de droit et à l'École supérieure d'études internationales de l'Université Laval. Coresponsable de l'axe «Arts, médias et diversité culturelle» de l'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique (OBVIA), elle s'intéresse plus particulièrement à la préservation de la diversité des expressions culturelles dans l'environnement numérique et au traitement des biens et des services culturels dans les accords de commerce.

d'IA, c'est-à-dire les données avec lesquelles on les entraîne. Un contrôle humain, qui peut intervenir sur les paramètres ou sur le contenu de ce que suggérerait un système d'IA, s'impose à **chacune des étapes clés** de la diffusion et de la distribution. En matière de diffusion, par exemple, une génération automatique de la grille d'office est certainement souhaitable, mais elle doit demeurer modifiable manuellement par les représentants en fonction des besoins exprimés par les détaillants. De la même manière, il faudrait évidemment pouvoir ajuster la grille d'office (par des prénotés ou des compléments d'office), car il est possible qu'un libraire souhaite obtenir une quantité moindre ou, au contraire, plus importante d'un ouvrage en raison de paramètres qui lui sont propres (comme la gestion des stocks ou un engouement local pour un titre en raison de la venue de l'auteur, etc.).

Les cas précis de Tamis et StoryFit

Tamis et StoryFit sont des compagnies qui utilisent l'IA pour l'analyse de textes, dont l'idée est principalement d'améliorer le référencement afin de bonifier la recommandation de titres à des lecteurs donnés. Derrière cet objectif, l'utilisation d'apprentissage machine et de reconnaissance émotionnelle permet de **classer les textes** en fonction de mots-clés (indiquant par exemple si le titre contient du suspense ou s'il est bien écrit) pour mieux en assurer la diffusion. Par ces applications, il devient possible de donner une seconde vie à certains ouvrages qui commencent à dater : en identifiant des mots-clés très précis sur ces ouvrages, la maison d'édition pourra efficacement en faire la **promotion** auprès du bon public ou au moment le plus opportun (si une thématique liée occupe à nouveau l'actualité dans les médias, par exemple). Certaines de ces compagnies peuvent même aller jusqu'à **automatiser**

Quelques chiffres

En utilisant des algorithmes d'IA pour recommander de façon individualisée des titres à ses lecteurs, l'éditeur allemand EbnerVerlag a enregistré des taux de conversion (autrement dit, de vente après publicité) près de vingt fois plus élevés qu'auparavant.

l'attribution des codes BISAC (nom d'un des codes de classement des livres), voire à **recommander** le nombre d'exemplaires à publier et à distribuer en fonction du contenu du titre et de son lectorat potentiel.

Synthèse

Si l'utilisation d'IA dans le secteur de la distribution du livre est une réalité depuis un certain nombre d'années, notamment chez les plus gros acteurs comme Amazon, le secteur de la diffusion a quant à lui encore une belle marge de progression pour l'introduction de cette technologie. Au-delà de l'établissement de corrélations auparavant inconnues entre certaines informations, l'IA peut servir à l'analyse en temps réel de données et ainsi aider à l'automatisation de certaines

prises de décision. Ici encore, la **concertation** entre les différents acteurs de la chaîne du livre quant à la question de l'accès et de la mise à disposition des données est certainement une piste à suivre.

Recommandations

- Des **accords de collecte et d'exploitation** des données et des métadonnées pourraient être passés entre les diffuseurs/distributeurs et les autres acteurs du livre, tant en amont (les éditeurs) qu'en aval (les détaillants, notamment les libraires), afin de créer des outils logiciels partagés et des retours sur investissement bénéficiant à plusieurs secteurs de la chaîne.
- Les protocoles d'échanges entre distributeurs, diffuseurs et clients (comme la norme NEEDA au Québec) doivent être **réévalués** au regard des nouveaux types de collecte de données visant à l'utilisation d'IA.
- Concernant plus précisément la distribution, les investissements financiers et humains pour la modernisation des espaces de stockage seront vraisemblablement très conséquents, puisqu'ils feront appel en grande partie à des équipements robotisés. Cet aspect doit être pris en compte, notamment par les **pouvoirs publics**, afin d'établir en partenariat des stratégies à moyen et long terme pour que les structures de petite et de moyenne taille ne soient pas laissées pour compte dans cette transformation.



Bonifier l'accès au livre

Si la transaction de livres peut avoir lieu en ligne ou dans la grande distribution, nous nous intéresserons ici à celle qui a lieu en librairie et à son pendant public, l'emprunt en bibliothèques. Car s'il est bien un point sur lequel la vente en ligne diffère de celle qui a lieu dans les librairies ou sur l'emprunt en bibliothèque, c'est bien celui du lieu et de la confiance qui lui est généralement attachée. Outre la valeur évidente de la proximité (et du stock immédiatement disponible sur les étagères), ces lieux de transaction du livre bénéficient d'une faveur particulière de la part des lecteurs – un **lien particulier**, fondé sur la confiance et la compétence, s'établit avec les libraires et les bibliothécaires. Dans un tel contexte, l'IA pourrait sembler de prime abord un instrument peu utile aux libraires et aux bibliothécaires, voire être perçu de manière négative. C'est une vision réductrice, car à titre d'aide à la recherche et à la recommandation, l'IA peut ouvrir bien des possibilités d'affaires dans un secteur largement en demande de solutions pérennes sur le plan économique.

En effet, et ce n'est pas une surprise, le moment de la vente ou de l'emprunt est celui qui épouse le plus naturellement les possibilités techniques de l'IA contemporaine, notamment au travers des systèmes de recommandation algorithmiques. Selon les chiffres de la Fédération des éditeurs européens (FEE), **610 000 nouveaux titres** ont été publiés dans la seule année 2017 dans l'espace européen. Un tel chiffre donne le tournis, car aucune librairie ni aucune bibliothèque ne peut stocker un tel nombre d'ouvrages, a fortiori employer un personnel à la mémoire suffisante pour les recommander. À de tels problèmes, l'IA peut apporter une ou plusieurs solutions. Et si la plupart des outils en la matière restent encore à développer (par le biais d'applications ou de logiciels professionnels), il est déjà possible de se faire une idée des utilisations potentielles de telles technologies. Un exemple :

Un lecteur entre dans une bibliothèque ou dans sa librairie préférée ; il part bientôt en vacances à Barcelone et aimerait emporter de la lecture, pourquoi pas un roman policier. Bien sûr, en l'état actuel des choses, tout bon bibliothécaire ou libraire fera avant tout appel à sa mémoire et à son expérience pour le conseiller. Demain, il pourra également se servir d'un logiciel propulsé par IA : grâce à un profil très détaillé des préférences du lecteur, le libraire/bibliothécaire de demain utilisera ce logiciel pour lui conseiller précisément un roman policier se déroulant à Barcelone (et, pourquoi pas, des extraits de textes en autoédition correspondant à son voyage). En poussant les choses plus loin, on peut imaginer de préciser encore les critères, par exemple pour que le roman corresponde également à un lectorat jeunesse (et que les enfants du client puissent le lire ensuite, par exemple).

Pour que de telles applications ou de tels logiciels se développent, le lien de confiance entre le lecteur et son libraire est primordial, et il en va de même en bibliothèque. C'est en fonction de cette confiance que le lecteur acceptera ou non de livrer ses préférences et qu'il bénéficiera demain de nouveaux outils de recommandation.

Prospectives

- Si le travail de référencement des livres a été correctement effectué, on peut alors imaginer qu'un système propulsé par IA permettra aux libraires et aux bibliothécaires de naviguer de façon très précise dans les différents catalogues existants en fonction des besoins et des envies des lecteurs potentiels (un livre avec beaucoup de suspense, qui se passe au Japon, etc.).

Quelques chiffres

Selon une étude de la Harvard Business Review, les compagnies faisant usage d'IA pour améliorer leurs ventes, tous domaines confondus, pouvaient augmenter leur nombre de clients potentiels de plus de 50% et réaliser des réductions de coûts de l'ordre de 40% à 60%.

- Le lien de confiance établi avec le libraire permettrait la création de bases de données **propres à chaque librairie**, tandis que les bibliothèques pourraient sans doute plus facilement les mutualiser. Les profils de lecteurs pourraient ainsi être affinés au fil des ans, non seulement en fonction des lectures et des titres achetés, mais également en fonction des conseils éclairés des libraires et des bibliothécaires. Cet affinage des recommandations (par les envies du lecteur, par le professionnel et par le système informatique) n'en serait que plus pertinent et complet, parce qu'évolutif.
- Il serait bien sûr possible d'envoyer directement aux lecteurs des **suggestions automatiques** de titres qui correspondent à leurs profils de lecteurs. Fondées sur la base de données établie à l'interne, ces recommandations automatisées n'en seraient que plus pertinentes.
- D'autres croisements de données sont à envisager : le catalogue d'un libraire pourrait, par exemple, inclure non seulement les ouvrages édités, mais aussi des ouvrages autoédités à commander grâce à de l'impression à la demande. Ouverte également sur l'autoédition recommandée par IA, la vente développerait un marché complémentaire, aidant à **promouvoir certains auteurs locaux** par exemple. C'est sans compter les croisements possibles avec les données d'autres secteurs culturels comme le cinéma ou le théâtre, qui ouvriraient sans doute sur des possibilités inédites de recommandation.
- Le catalogue des titres d'une librairie ou d'une bibliothèque pourrait faire l'objet d'une évaluation en temps réel, de sorte de générer la liste des commandes à faire, en planifier d'autres ou encore indiquer quels titres ne se vendent ou ne se louent pas pour suggérer un retrait des tablettes ou (dans les pays où ne s'applique pas la loi sur le prix unique du livre) proposer des réductions sur ces derniers.

Les besoins

Sans surprise, le logiciel qui assistera le professionnel du livre dans ses recommandations a encore besoin de données. En plus de celles, détaillées, sur les livres en vente ou dans le catalogue de l'institution, il faudrait colliger et croiser les informations sur chacun des lecteurs de la librairie ou de la bibliothèque. En isolant les informations d'ordre personnel (celles qui permettent d'identifier

la personne et qui sont soumises à un certain nombre de règles, notamment celles qui sont liées à la législation sur la protection des données personnelles), les analyses proposées auraient pour intérêt d'approfondir le profil des lecteurs et leurs préférences de lecture : la taille des livres souhaités, le genre préféré, l'envie de découvrir autre chose ou au contraire de rester dans tel ou tel champ littéraire, l'envie de lire davantage de livres avec une dimension féministe, etc. Ces données seraient simplement récoltées à l'aide d'un questionnaire intégré dans un logiciel à destination des professionnels, qui pourraient en gérer l'utilisation en accord avec les lecteurs.

Quelques chiffres

D'après l'agence américaine de distribution des codes ISBN Bowker, plus de 1 680 000 codes ISBN ont été distribués pour l'année 2018, une augmentation de 40% par rapport à 2017. Ce secteur extrêmement dynamique n'est pas particulièrement répertorié ni exploité par les libraires – surtout en raison des cadres légaux et réglementaires –, ce que les technologies d'IA pourraient pourtant faciliter.

Les limites

Les outils de recommandation propulsés par IA n'ont pas vocation de remplacer les libraires ou les bibliothécaires, mais de les assister en leur permettant, par exemple, de naviguer dans de plus grands catalogues tout en affinant et en organisant leur connaissance de leur clientèle. Pour autant, ces recommandations – de même que les « profils » de lecteurs – ne sont que des **modèles** : ils ne représentent que des versions partielles, limitées des goûts des individus et des titres répertoriés. Si ce n'était pas le cas, les recommandations de livres vendus sur les plus grands sites de commerce en ligne feraient

mouche à chaque fois, ce qui n'est pas le cas. Un système de recommandation par IA n'est donc pas parfait, ce pour quoi les métiers de libraire et de bibliothécaire auront toujours leur pertinence.

Il n'en demeure pas moins qu'avec l'IA, il existe également un espace à investir pour les professionnels du livre, un espace qui passe avant tout par la **confiance** que leur offre leur clientèle. Cette confiance gagne à être préservée et à être valorisée, notamment en respectant les règles en vigueur en matière de fichiers clients et de données personnelles. Cette approche « augmentée » de la recommandation est sans doute **l'arme secrète** des libraires et des bibliothécaires face à la concurrence tout algorithmique des géants de la vente en ligne.

Le cas de BookGenie, de HarperCollins

HarperCollins a intégré en 2017 un outil de recommandation de livres sur Facebook, BookGenie. Le lecteur qui souhaitait obtenir des recommandations sur ses prochaines lectures pouvait ainsi cliquer sur l'onglet « Écrire un message » de la page Facebook de l'éditeur et entrer en communication avec un chatbot. Disponible uniquement sur la page en anglais de l'éditeur, ce système proposait de pallier un système de recommandation qui ne serait basé que sur les anciens titres lus. De fait, le chatbot était interactif et suggérait au client potentiel des titres qui prenaient en compte différents facteurs, dont son humeur du moment et ses goûts de lecteur. Bien évidemment, ce système ne recommandait que les titres de HarperCollins et pas ceux de maisons d'édition concurrentes.

L'expertise de Clément Laberge

Contrairement à d'autres types de contenus culturels, les livres restent encore difficilement lisibles par des machines. Leur contenu demeure le plus souvent inaccessible aux algorithmes qui tissent des liens entre tous les types d'information qui se trouvent dans le web – et ne peut donc pas servir à améliorer leur découvrabilité. Dans ce contexte, je pense qu'il faut surtout voir l'IA comme un puissant moyen d'offrir aux livres la place qu'ils méritent dans cette grande toile du savoir et de l'imagination humaine. L'IA doit servir à outiller les libraires et les bibliothécaires pour qu'ils accompagnent encore plus efficacement les lecteurs à travers tout ce qui s'écrit – qui ne prend plus toujours la forme traditionnelle du livre et qui n'est plus forcément distribué par les canaux commerciaux habituels. Les bibliothèques ont un rôle stratégique fondamentale à jouer dans le développement de l'IA dans le monde du livre – notamment grâce aux impressionnantes quantités de données bibliographiques et de statistiques d'emprunt dont elles disposent.

Ces données devraient autant que possible être partagées avec l'ensemble des acteurs locaux du monde du livre, à défaut de quoi l'IA restera l'apanage de quelques géants du Web et répondra essentiellement à des impératifs commerciaux. Pour cette raison, les bibliothèques devraient constituer la pierre d'assise de toute stratégie d'IA pour le monde du livre.

Clément Laberge a travaillé pour plusieurs entreprises dans le secteur de l'édition au Québec et en France. Il a exercé des rôles importants dans plusieurs grands chantiers de diffusion numérique. Aujourd'hui consultant indépendant, il s'intéresse particulièrement aux enjeux de distribution et de découvrabilité de tous les types de contenus culturels numériques.



Synthèse

En matière de vente, l'IA peut assister les libraires et les bibliothécaires en leur permettant de se retrouver dans des catalogues de plus en plus imposants, voire en leur ouvrant de nouveaux marchés, par exemple en matière d'autoédition. On serait, là encore, plus avisés de parler d'intelligence augmentée que d'intelligence artificielle. La nécessaire acquisition de données peut paraître écrasante pour certains, d'autant que l'on sait que les plus gros joueurs, Amazon en tête, disposent d'une longueur d'avance sur ces questions. C'est cependant oublier que les bibliothèques disposent depuis déjà un certain nombre d'années de données pertinentes sur les lecteurs – par leurs statistiques d'emprunt de livres papier et de livres numériques –, données qui peuvent constituer un avantage concurrentiel considérable. De la même manière, les librairies ont un avantage comparatif indéniable sur les entreprises uniquement en ligne, car elles peuvent contre-interroger leur lectorat, obtenant un retour plus précis qu'un simple nombre d'étoiles sur les titres lus (par exemple : j'aime le début mais pas la fin, il me semble que l'histoire est peu crédible, ce livre me rappelle tel autre, etc.). Autant de données qui, si elles sont collectées, constituent des actifs auxquels la concurrence n'a pas accès. Ces données ont conséquemment une grande valeur pour le reste de l'écosystème du livre, que ce soit pour les éditeurs, les diffuseurs ou même les auteurs.

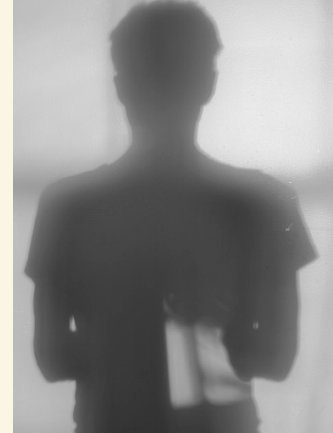
Recommandations

- Quels qu'ils soient, les outils destinés à améliorer la recommandation et les ventes grâce à l'IA ne peuvent être pertinents que s'ils résultent d'une **concertation** avec les éditeurs, lesquels sont dans la meilleure position pour produire des analyses de données de leurs titres. En retour, les librairies et les bibliothèques sont dans une position privilégiée pour la récolte de données sur leurs lecteurs. Logiquement, ces différents jeux de données constitueront une richesse d'autant plus importante s'ils sont mutualisés entre les différents acteurs, qu'ils soient des joueurs indépendants, des institutions publiques ou des organisations impliquées dans la circulation des données sur le livre.
- Une réflexion entre les acteurs publics et les bibliothèques doit être mise en place à l'égard des données que ces dernières possèdent. En effet, l'avantage concurrentiel conséquent que constituent ces ressources a tout intérêt à être mise à la disposition du monde du livre, que ce soit au travers d'accords sectoriels particuliers ou sous forme de mutualisation.



« Prospectives littéraires, par David «Jhave» Johnston

De nos jours, l'acte d'écrire de la littérature demeure un métier de l'esprit, du cœur et de l'écoute. Mais on peut supposer que dans un proche avenir littéraire, beaucoup de choses changeront (comme les technologies liées à l'écriture) et beaucoup de choses resteront les mêmes (comme l'esprit, le corps, la société; comme l'amour, la mort, la langue). Il est maintenant possible d'imaginer une fonction d'autocomplétion qui serait configurable par l'utilisateur. Cette fonction pourrait mettre à jour ses modèles d'apprentissage profond en temps réel, modèles formés sur les mots de l'écrivain (avec peu ou pas besoin de données). L'écrivain sélectionnerait les genres, les styles, les idiomes, les idiosyncrasies régionales et le niveau de complexité de la langue à partir d'une carte de classification dynamique et évolutive, puis un modèle d'autocomplétion personnalisé (l'assistant archétypal) générerait un texte cohérent. D'autres chemins littéraires apparaîtraient alors sous forme de tentacules ramifiés. Alors que les écrivains débutants et les enfants apprennent à lire et à écrire au travers de canaux de rétroaction en temps réel, les écrivains adultes deviennent des curateurs éditoriaux, éliminant une avalanche de contenu généré par intelligence artificielle (IA). Le danger est le même qu'il l'a toujours été: conformité, prévisibilité, réplique obsolète, boucles de renforcement du contenu. L'intérêt est de bénéficier d'une créativité toujours présente, la muse infinie de l'IA, ouvrant la porte à des auteurs hybrides homme-machine, à des genres littéraires IA-humains – des espaces d'écriture semi-automatiques ouverts et évolutifs. Pourtant, comme je le dis souvent, l'intelligence artificielle contemporaine (principalement les réseaux de neurones) est faible en ce qui concerne la subtilité d'un contexte, le flux narratif, l'acuité psychologique, la continuité. Une IA supérieure et nouvelle, plus fine, dérivée de l'architecture neurologique et incorporant des éléments d'expériences vécues, sera nécessaire. Ainsi, cette IA du futur pourrait bien éclipser tous les auteurs, hormis les plus forts, tant en termes de polyvalence que de rapidité. Les cerveaux de chair deviendraient alors de simples données, et les algorithmes autonomes d'IA pourraient écrire une élégie (ou des nouvelles sportives, des romans populaires, des potins, la météo, des nouvelles) à l'intention de la terre ravagée, de ses créatures subjuguées et ses écosystèmes en déclin.



David «Jhave» Johnston est un poète, vidéaste et artiste canadien qui travaille essentiellement sur les médias numériques. Il a enseigné de 2014 à 2017 à la School of Creative Media de l'Université de Hong Kong, avant de revenir s'installer à Montréal. Il récemment publié la somme de poèmes cocréés par IA *ReRites* chez Anteism Press (2019).



Conclusion

L'intelligence artificielle dans le monde du livre est une réalité. Des acteurs comme Amazon y investissent massivement depuis des années, comme ils ont rapidement vu la valeur de leurs propres données, et continueront certainement d'y placer leurs jetons dans le futur. Face à cette concurrence, chacun des acteurs de la chaîne du livre doit se demander comment faire face à ce mouvement de fond, qui s'apprête à modifier en profondeur de nombreux métiers du livre. Pour s'adapter, il faut garder en tête que les technologies d'IA seront d'autant plus pertinentes et efficaces qu'elles auront été réfléchies en amont de l'écosystème dans lequel on les utilise.

En ce sens, la voie dorée pour l'introduction d'IA dans les différents maillons de la chaîne est sans aucun doute celle d'une **exploitation des différentes données déjà disponibles et que ne possède pas la concurrence**, quitte à nouer des accords asymétriques entre les différents acteurs intéressés pour y accéder. Ces données existent, ce sont des actifs précieux que tous les acteurs du milieu du livre possèdent, particulièrement dans les bibliothèques ; elles constituent en tant que telles des ressources stratégiques considérables.

Leur mise à profit, tout comme les initiatives d'introductions technologiques ou de mutualisations éventuelles des ressources, font bien sûr face à de nombreux obstacles, que ce soit la force d'inertie des usages, le coût (en planification et en argent) de leur implantation ou l'état du droit et des règlements. Ces cadres ne doivent pas être minimisés : différents pour chaque pays, ils peuvent cependant – si les acteurs publics concernés se mobilisent sur la question – servir de leviers éventuels pour la mise en place de ces solutions d'IA efficaces et pertinentes pour le monde du livre. À terme, la mise en réseau des informations collectées participerait idéalement d'une synergie de transmission de l'information tout à fait bénéfique pour l'ensemble de l'écosystème.

De façon à bien la planifier et à l'arrimer étroitement avec les ressources existantes, toute stratégie d'adoption de l'IA doit prendre en considération les trois volets suivants :

I. L'ACCESSIBILITÉ DES DONNÉES	
Évaluer ses données déjà disponibles	Le meilleur avantage en matière d'IA, c'est la donnée que la concurrence ne possède pas. Sur ce point, l'écosystème du livre dispose notamment d'une ressource stratégique inaccessible aux seuls acteurs numériques : les statistiques centralisées des distributeurs, les données des éditeurs ou encore celles collectées par les bibliothèques et les libraires, qu'il importe de mettre à plat et d'apprendre à bien connaître.
Établir des collaborations pertinentes	Chaque acteur de la chaîne du livre dispose de données que lui seul possède et il a besoin de données qu'il ne possède pas. Dans ce contexte, nouer des accords avec les joueurs les plus pertinents pour son modèle d'affaires afin de diversifier ses ressources en données peut constituer une stratégie payante.
Augmenter la qualité et le volume des données collectées	Le fonctionnement des applications d'IA étant lié au volume et à la qualité des données disponibles, il importe de planifier son scénario optimal de collecte en découvrant quels ensembles de données ne sont pas moissonnés, en définissant les renseignements utiles à récolter, ainsi qu'en établissant la stratégie pour y parvenir.

S'assurer de solutions fiables et sécuritaires pour le stockage des données	La sécurité des informations est primordiale pour tous, mais il existe différentes façons de l'assurer. Une grande maison d'édition ne souhaitera sans doute pas conserver ses données sur un serveur externe, tandis qu'une petite librairie aura peut-être tout intérêt à en mutualiser la conservation en ligne pour des raisons de coûts.
Disposer de données uniformisées et facilement exportables	De la compatibilité des données entre elles dépend le succès des solutions d'IA qui en tireront parti. Sur cette question fondamentale, établir un protocole de normalisation des données en accord avec tous les acteurs concernés est un préalable nécessaire, sur lequel les pouvoirs publics ont un rôle à jouer.
2. L'UTILISATION DES DONNÉES	
Répertorier les problèmes pouvant être réglés par IA	Un audit de son secteur (macro) et de son entreprise en particulier (micro) doit être mené afin d'évaluer quelles décisions peuvent être automatisées, quels processus peuvent être optimisés, quelles informations stratégiques pourraient révéler des corrélations pertinentes, etc. Cette étape doit conduire à l'établissement d'objectifs d'intégration opérationnelle de ces technologies.
Catégoriser et étiqueter les données en fonction des objectifs visés	Les données collectées sont à peu près inutiles sans cette étape fondamentale de préparation des données. Leur catégorisation et leur étiquetage doivent s'inscrire dans un objectif précis (lequel peut être intersectoriel, interentreprises ou peut n'intéresser qu'un acteur précis).
3. LA MISE À PROFIT DES DONNÉES	
Définir des objectifs de développement prioritaires	En prenant en compte la disponibilité de ses données et ses besoins les plus manifestes, établir sa feuille de route des développements à prioriser permettra d'organiser sa stratégie de façon cohérente en l'accordant notamment à la collecte de données et à leur uniformisation conséquente.
Impliquer les instances pertinentes	Afin d'orienter les chantiers communs et de contribuer au financement du développement de prototypes ou d'éléments logiciels partagés, des institutions et des acteurs gouvernementaux devront être mobilisés. La mise en commun de données appelle une concertation, notamment dans le développement de ressources logicielles (aussi rendu possible par la création d'un cadre réglementaire favorable).
Identifier les solutions d'IA pertinentes	L'IA étant une constellation de technologies différentes, des expertises informatiques seront requises à ce stade pour l'évaluation et la mise en place des solutions d'IA, lesquelles peuvent être très larges et requérir un développement informatique spécifique ou au contraire bénéficier en partie de solutions préexistantes.

Bibliographie

- Sofian Audry, «*for the sleepers in that quiet earth.*: Experiencing the Behavior of a Deep Learning Neural Network Agent through a Generative Artbook», *Proceedings of the 24th International Symposium on Electronic Art*, 2018, p. 270-277.
- Alexandra Bensamoun & Joëlle Farchy, rapport de mission *Intelligence artificielle et culture*, pour le compte du Conseil Supérieur de la propriété littéraire et artistique, 2020, 106 p., <https://www.culture.gouv.fr/Sites-thematiques/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux/Missions/Mission-du-CSPLA-sur-les-enjeux-juridiques-et-economiques-de-l-intelligence-artificielle-dans-les-secteurs-de-la-creation-culturelle>
- Margaret Boden, *AI: Its nature and future*, Oxford University Press, 2016, 208 p.
- Annemarie Bridy, «Coding creativity: copyright and the artificially intelligent author», dans *Stan. Tech. Law Review*, 2012, 28 p.
- Dominique Cardon, Jean-Philippe Cointet et Antoine Mazières, «La revanche des neurones – L'intention des machines inductives et la controverse de l'intelligence artificielle», dans *Réseaux*, vol. 5, n° 211, 2018, p. 173-220.
- Matt Carlson, «The robotic reporter. Automated journalism and the redefinition of labor, compositional forms, and journalistic authority», *Digital Journalism*, vol. 3, n° 3, 2015, p. 416-431.
- Nathalie Casemajor, Guy Bellavance, Guillaume Sirois *et al.*, «Pratiques culturelles numériques et plateformes participatives : opportunités, défis et enjeux», rapport de recherche, FRQSC, 2019, 192 p., <https://tinyurl.com/y87krndu>
- Grégory Chatonsky & Karmel Allison, *Machines Upon Every Flower*, Anteism Press, 2018, 64 p., <https://www.anteism.com/shop/machines-upon-every-flower>
- Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, 2015, 352 p.
- Ross Goodwin, *1 the Road*, [Français, Anglais], Jean Boîte Éditions, 2018, 143 p.
- Gould Finch & Frankfurter Buchmesse, «The Future Impact of Artificial Intelligence on the Publishing Industry», *Livre blanc*, 2019, 24 p.
- Tristan Greene, «How AI will change storytelling and become the next great art form», *The Next Web*, 2019, <https://thenextweb.com/artificial-intelligence/2019/01/25/how-ai-will-change-storytelling-and-become-the-next-great-art-form/>
- Andrea Guzman, «What Is Human-Machine Communication, Anyway?», *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves*, Peter Lang, 2018, p. 1-28.
- Ian Heller, Michael Wu et Austin Garrison, «The Future of AI in Distribution. Artificial Intelligence, Machine Learning and the Implications for Distributors», MDM/Pros, 2019, 18 p., <https://tinyurl.com/y73emhje>
- Véronique Heurtematte, «Intelligence artificielle, une nouvelle donne pour les bibliothèques universitaires», *Livres Hebdo*, 2019, <https://www.livreshebdo.fr/article/intelligence-artificielle-une-nouvelle-donne-pour-les-bibliotheques-universitaires>
- Hervé Hugueny, «Enquête sur l'intelligence artificielle dans l'édition», *Livres Hebdo*, 2018, <https://www.livreshebdo.fr/article/enquete-sur-lintelligence-artificielle-dans-ledition>
- Hervé Hugueny, «Plongée dans l'intelligence artificielle appliquée au livre», *Livres Hebdo*, 2017, <https://www.livreshebdo.fr/article/plongee-dans-lintelligence-artificielle-appliquee-au-livre>
- David «Jhave» Johnston, *Raw Output / Responses (ReRites)*, Anteism Press, 2019, 92 p., <https://www.anteism.com/shop/rerites-david-jhave-johnston>
- Howard Lovy, «Artificial Intelligence and the Indie Author: AskALLi Advanced Self-Publishing Salon with Orna Ross and Joanna Penn», *Alliance of Independent Authors*, 2019, <https://selfpublishingadvice.org/artificial-intelligence/>
- Francesco Marconi, Alex Siegman et Machine Journalist, «The Future of Augmented Journalism: A Guide for Newsrooms in the Age of Smart Machines», Associated Press, 2017, https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf
- Nils J. Nilsson, *The Quest for Artificial Intelligence, A History of Ideas and Achievements*, Cambridge University Press, 2010, 580 p.
- Novipro & IBM, *Intelligence Artificielle, Comment débiter un projet d'IA dans votre entreprise*, Livre blanc, 16 p., <https://tinyurl.com/y8dqq7zu>
- Allison Parrish, *Articulations*, Counterpath Press, 2018, 120 p.
- Holly Lynn Payne, «Why Book Publishing Seeks Artificial Intelligence», *HuffPost*, 2017, https://www.huffpost.com/entry/why-book-publishing-seeks-artificial-intelligence_b_59482841e4b0961faacbe5d6
- Gregorio Pellegrino, «Improving automatic image description in EPUB using Artificial Intelligence», 2019, <https://www.youtube.com/watch?v=XZpgGNoBQoo>

- Adam Rowe, « Interactive Storytelling App Novel Effect Just Raised A \$3M Series A », *Forbes*, 2018, <https://www.forbes.com/sites/adamrowe1/2018/05/22/storytelling-app-novel-effect-3m-series-a/#1a5c0f1a120d>
- SAP, « Trends in the Wholesale Distribution Industry. Riding the Waves of Disruption in Wholesale Distribution. Adapting Business-to-Business Paradigms to the Experience Economy », *Livre blanc*, 2019, 12 p.
- Satoshi Sato, « A challenge to the third Hoshi Shinichi award », *Proceedings of the INLG 2016 Workshop on Computational Creativity in Natural Language Generation*, 2016, p. 31-35.
- Paul Sawers, « Huawei's StorySign app can translate kids' books into sign language », *Venturebeat*, 2018, <https://venturebeat.com/2018/12/03/huaweis-storysign-app-can-translate-kids-books-into-sign-language/>
- Robin Sloan, « How to Write a Novel with Machine Learning? », *Ars Live #25*, Entrevue avec Ars Technica, 2018, <https://video.arstechnica.com/watch/ars-live-25-how-to-write-a-novel-with-machine-learning>
- SNE [Syndicat National de l'Édition], *Les Chiffres de l'édition, rapport statistique du SNE 2018-2019*, 24 p., <https://tinyurl.com/y9thfejp>
- Pierre Talbot, *Livre blanc LA*, CCI Essonne, 2019, 20 p.
- Vincy Thomas, « HarperCollins mise sur l'intelligence artificielle », *Livres Hebdo*, 2017, <https://www.livreshebdo.fr/article/harpercollins-mise-sur-lintelligence-artificielle>
- Deb Vanasse, « Publishing by the Numbers », *IBPA Independant*, 2017, <https://articles.ibpa-online.org/article/publishing-by-the-numbers/>
- Boris Vian, « Un robot-poète ne nous fait pas peur », dans *Œuvres romanesques complètes*, Tome II, Paris, Gallimard (Bibliothèque de la Pléiade), 2010, p. 1021-1025.
- Sophie Viaris de Lesegno, « Les contrats de diffusion-distribution », *LEGICOM*, vol. 24, n° 1, 2001, p. 29-36, <https://www.cairn.info/revue-legicom-2001-1-page-29.htm>
- [Sans auteur], « Publishing in the Era of Big Data », *Kobo Whitepaper*, 2014, <https://www.yumpu.com/en/document/read/42663323/publishing-in-the-era-of-big-data-kobo-whitepaper-fall-2014>

