
La génération automatique de textes en presse écrite : historique problématisé, questions d'éthique et analyse de contenu

Auteur : Schevenels, Hanna

Promoteur(s) : Vanesse, Marc

Faculté : Faculté de Philosophie et Lettres

Diplôme : Master en journalisme, à finalité spécialisée en investigation multimédia

Année académique : 2019-2020

URI/URL : <http://hdl.handle.net/2268.2/9403>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Université de Liège
Faculté de Philosophie et Lettres
Département Médias, Culture et Communication



**LA GÉNÉRATION AUTOMATIQUE DE TEXTES
EN PRESSE ÉCRITE :**

HISTORIQUE PROBLÉMATISÉ, QUESTIONS D'ÉTHIQUE ET
ANALYSE DE CONTENU

Mémoire présenté par SCHEVENELS Hanna
en vue de l'obtention du grade de
Master en journalisme à finalité spécialisée en
investigation multimédia

Année académique 2019-2020

REMERCIEMENTS

Je tiens tout d'abord à remercier Monsieur Marc Vanesse, Directeur de ce travail de fin d'études, qui a accepté de me suivre dans la concrétisation de ce projet académique qui marque la fin de mon cursus universitaire. J'exprime également toute ma gratitude envers Monsieur Boris Krywicki, assistant chercheur, dont les remarques m'ont grandement aidée à préciser mon domaine d'analyse. Je les remercie tous deux de m'avoir accompagnée, encouragée et conseillée avec réactivité tout au long de cette entreprise, aussi challengeante qu'enrichissante.

Je suis également très reconnaissante envers Monsieur Alain Jenotte, professeur de datajournalisme, pour sa disponibilité et ses conseils éclairés. Son cours de « Techniques Numériques du Journalisme » a joué un rôle crucial dans le choix et la définition de mon domaine de recherche. Je le remercie, à travers ce cours, d'avoir éveillé ma curiosité et de m'avoir donné l'occasion de découvrir un univers qui m'était jusqu'alors inconnu.

Je souhaite aussi remercier mon compagnon, Alexandre Lhoest, étudiant en ingénieur *Data Science* à l'Université de Liège, dont l'intérêt pour l'informatique et l'intelligence artificielle m'a assurément encouragée à me diriger vers ces recherches captivantes, mêlant sa passion et la mienne. J'adresse également une pensée particulière à mes proches, ma famille et plus particulièrement à mes parents, dont l'affection et le soutien sans faille m'ont permis de mener à bien ce projet, ainsi que tous ceux qui ont ponctué mon parcours universitaire. Je leur dédie modestement ce mémoire et je ne les remercierai jamais assez pour l'aide qu'ils m'ont apportée tout au long de cette aventure.

Enfin, merci à Marie Guérin, Émilie Mascia, Nadia Chater, Ricardo Gutiérrez, André Thiel, Xavier Ducarme, Rodolphe Magis, David Coppi, Julien Covolo, Nicola Becquet, Laurence Dierickx et Claude de Loupy, qui ont accepté de répondre aux questions et interviews qui ont ponctué ces recherches passionnantes.

Hanna Schevenels

SOMMAIRE

REMERCIEMENTS	1
SOMMAIRE	2
GLOSSAIRE	3
INTRODUCTION	7
CHAPITRE 1 : Historique problématisé.....	16
CHAPITRE 2 : Questions clés et enjeux	52
CHAPITRE 3 : Étude de contenu.....	85
CONCLUSION.....	96
BIBLIOGRAPHIE	101
ANNEXES	114
TABLE DES FIGURES	145
TABLE DES MATIÈRES.....	146

GLOSSAIRE

IA : Acronyme d'intelligence artificielle¹. Le terme a été conçu par le mathématicien américain John McCarthy en 1955 (Latar, 2018). L'IA est souvent définie comme la science et l'ingénierie dont le but est de rendre des machines intelligentes, ou encore de les faire réfléchir et agir comme des humains (Louppe, 2017).

IA FAIBLE : Il s'agit du seul type d'IA développé jusqu'alors. L'IA faible est caractérisée par l'impossibilité de réaliser plusieurs tâches simultanément. Les systèmes d'IA faible sont notamment capables de reconnaître une image, de trouver des données dans des *data silos*², de répondre aux questions d'utilisateurs, de réaliser des traductions ou encore de jouer aux échecs. Tous les systèmes d'IA existants sont considérés comme des IA faibles (Dickson, 2017).

IA FORTE : Les partisans de la théorie de l'IA forte sont convaincus qu'il sera, un jour, possible de créer un être artificiel doté d'une conscience, d'un esprit, ou encore d'une sensibilité. Ce stade n'a encore jamais été atteint ; il n'existe aujourd'hui aucun système d'IA forte (Dickson, 2017). Ray Kurzweil, ingénieur et futurologue au sein de la société américaine Google, est considéré comme l'un des adhérents les plus éminents à l'existence possible de ce type d'IA (Levy, 2013).

ALGORITHMES INTELLIGENTS : Contrairement aux algorithmes ordinaires, qui sont des ensembles d'instructions que les ordinateurs exécutent dans un ordre prédéterminé, les algorithmes intelligents sont programmés de façon à assigner des probabilités à des événements spécifiques. Ils apprennent seuls de leurs expériences, réduisent graduellement les niveaux d'incertitudes au cours du temps et se créent, par conséquent, de nouvelles connaissances. Alors que les algorithmes ordinaires sont programmés pour réaliser les mêmes actions inlassablement, les algorithmes intelligents modifient ainsi le cours de leurs opérations suite à un processus d'apprentissage. La principale force des algorithmes intelligents réside dans leur capacité à opérer au cours de situations très complexes, en scannant d'énormes *data silos* à différentes variables et à très grande vitesse. Grâce aux

¹ Dans ce travail, l'expression « intelligence artificielle » sera remplacée par son acronyme « IA ».

² Voir glossaire (p.4).

technologies informatiques actuelles, les algorithmes intelligents peuvent effectivement analyser et identifier des connexions statistiquement importantes parmi des bases de données de taille pratiquement illimitée (Latar, 2018).

BIG DATA (DONNÉES DE MASSE) : La notion de *Big Data* s'est développée parallèlement à la croissance démultipliée des données numériques, rendue possible grâce à la démocratisation d'Internet (Collmann & Matei, 2016). De façon générale, *Big Data* désigne des masses de données qui peuvent prendre différentes formes et qui sont générées dans de multiples domaines (Stone, 2014). Dans un contexte exclusivement économique, *Big Data* réfère aux données créées par les utilisateurs au cours de leur navigation sur le World Wide Web. Ces dernières peuvent être analysées afin d'améliorer les services des plateformes en question. Dans le cas des sites de presse, elles permettent notamment de personnaliser leurs différents contenus en fonction des utilisateurs. D'un point de vue documentaire, *Big Data* renvoie aux innombrables chiffres, informations et données disponibles sur le World Wide Web et au sein de bases de données publiques et privées (Bremme, n.d). Ce type de données de masse peut notamment être utilisé par les IA génératrices de textes pour concevoir des articles de presse automatiquement (Latar, 2018).

DATA SCIENCE (SCIENCE DES DONNÉES) : Science (pratiquée par les *data scientists*) qui a pour but d'extraire des informations utiles au cœur de données de masse issues du World Wide Web et de bases de données publiques et privées (Boyadjian, Olivesi, & Velcin, 2017). Elle privilégie une technique proche de la fouille de données : « extraction, stockage, indexation, analyse et catégorisation, puis visualisation » (Boyadjian et al., 2017, p.11).

DATA SILO : Désigne un ensemble de données isolé d'autres ensembles de données (TechTarget, 2015).

DEEP LEARNING (APPRENTISSAGE PROFOND) : Domaine relativement récent des recherches en IA dont le but est de créer un cerveau artificiel humain qui se rapproche le plus possible de celui du cerveau humain. L'invention de *Deep Learning* a mis fin à ce que beaucoup de scientifiques ont appelé « l'hiver de l'intelligence artificielle » (Hof, 2013). Dès 1986, ses pères fondateurs, Geoffrey Hinton, David Rumelhart et Ronald Williams, en ont établi les prémices (Somers, 2017). De façon générale, *Deep Learning* ajoute de la

profondeur et des strates d'analyse aux études réalisées dans le cadre des réseaux neuronaux³. Ce domaine de recherche est principalement basé sur le concept de *back propagation* (propagation en arrière), qui repose sur la logique suivante : lorsqu'un réseau neuronal tente d'identifier un objet donné (une phrase, par exemple), il en fait une analyse à partir des différentes couches le composant jusqu'à atteindre une décision binaire finale à son propos. Si l'identification est incorrecte, le système revient sur son analyse et corrige les erreurs faites à chaque étape. Le processus de *back propagation* permet notamment au système d'apprendre de ses erreurs de façon similaire à celle du cerveau humain. Ce domaine de recherche a notamment permis de réaliser des progrès remarquables concernant la reconnaissance vocale et la reconnaissance d'images. Le problème majeur lié au *Deep Learning* concerne la prise de décision des systèmes et plus précisément l'incapacité des développeurs à expliquer les décisions prises par les algorithmes (Latar, 2018).

GAFAM : Acronyme qui désigne les cinq multinationales les plus influentes sur le marché du numérique : Google, Apple, Facebook, Amazon et Microsoft. Ces dernières régissent le cœur informationnel de nos sociétés (Smyrniotis, 2016).

MACHINE LEARNING (APPRENTISSAGE AUTOMATIQUE) : Domaine de recherche dont le *Deep Learning* est une sous-catégorie. *Machine Learning* repose sur des méthodes d'apprentissage automatique des systèmes d'IA, leur permettant d'améliorer leurs fonctions suite aux erreurs faites précédemment (Benbouzid & Cardon, 2018).

NATURAL LANGUAGE PROCESSING (NLP) : Domaine de recherche interdisciplinaire à l'intersection entre l'IA, l'informatique, la science cognitive, le traitement de l'information et la linguistique. Le but ultime du NLP est de permettre aux systèmes d'IA de traiter et de comprendre le langage humain de façon intelligente. Les algorithmes de NLP ont déjà envahi de nombreux domaines de la vie courante et ont commencé à réorganiser la manière dont nous identifions, obtenons et utilisons certaines informations et données (Deng & Liu, 2018).

NATURAL LANGUAGE GENERATION (NLG) : Sous-discipline de la linguistique computationnelle dont le but est de transformer des données en un texte lisible,

³ Voir glossaire (p.6).

grammaticalement et syntaxiquement correct, sans intervention humaine (Reiter & Dale, 2000).

RÉSEAUX NEURONAUX : Algorithmes qui imitent le processus de traitement d'informations des neurones du cerveau humain. Ils fonctionnent au cœur d'un système neuronal artificiel et dépendent des comportements passés de ce dernier, ainsi que de la probabilité de connexions entre différents éléments. Le neurone artificiel est mis à jour à chaque fois qu'il rencontre de nouvelles informations ; il ne les transmettra aux autres neurones que si la somme totale de ces informations n'excède un certain palier. Ce raisonnement est caractéristique de la pensée rationnelle humaine (Latar, 2018).

INTRODUCTION

L'intelligence artificielle, l'humain de demain ?

1. Mise en contexte

Lors du second tour dans la région Auvergne-Rhône-Alpes, trois listes s'affrontaient pour les élections régionales 2015 : la liste Union de la Droite de M. Laurent WAUQUIEZ, la liste Union de la Gauche de M. Jean-Jack QUEYRANNE et la liste Front National de M. Christophe BOUDOT.

Les résultats du second tour à Savas sont les suivants : la liste Union de la Droite conduite par M. Laurent WAUQUIEZ était en tête des suffrages avec 45,81 %. Les autres listes (liste Union de la Gauche) et liste Front National) conduites par M. Jean-Jack QUEYRANNE et M. Christophe BOUDOT ont obtenu respectivement 28,33 % et 25,86 % des voix.

Dans cette localité, le taux d'abstention a atteint 40,03 %⁴.

Ces quelques lignes ont été rédigées par un robot. L'article repris ci-dessus, publié sur le site internet du journal *Le Monde*, a effectivement été généré sans aucune intervention humaine par un logiciel de génération automatique de textes appartenant à la société française Syllabs. Lors des dernières élections départementales françaises de 2015, plus de 32 000 articles ont ainsi été publiés sur le site internet lemonde.fr afin d'informer les lecteurs des résultats de votes de chacune des communes françaises (Bronner, 2015). Il s'agissait là d'une première mondiale, aucun autre média de langue française n'avait encore fait appel aux services d'une IA génératrice automatique de textes. Plus impressionnant encore, aucun logiciel d'IA n'avait jamais été programmé pour réaliser un travail de rédaction journalistique d'une telle envergure à partir d'un ensemble de données factuelles et chiffrées (Glad, 2015). Cette innovation du *Monde*, réalisée en partenariat avec la société Syllabs, s'inscrit au cœur d'un mouvement plus large d'utilisation de l'IA en journalisme. Plus précisément, cette performance témoigne d'une nouvelle tendance impliquant la génération automatique de textes en presse écrite (Diakopoulos & Koliska, 2017).

Aujourd'hui, les *news room* se trouvent à un tournant important. Selon Amir Ruskin (2018), à l'ère de *Big Data*⁵, les anciennes pratiques journalistiques ne se révèlent plus capables de maintenir l'intérêt des consommateurs. C'est dans ce contexte de crise que les générateurs

⁴ Data2Content (2015). Résultats du second tour des élections régionales : commune de Savas. *Le Monde*. Retrieved 27 March 2020 from

<https://abo.lemonde.fr/auvergne-rhone-alpes/ardeche,07/savas,07310/elections/regionales-2015/>

⁵ Voir glossaire (p.4).

automatiques de textes ont fait leur entrée dans le domaine de la presse écrite. Ils reposent sur un principe général d'entrées-sorties, un logiciel recevant en entrée (*input*) des données et produisant en sortie (*output*) un texte lisible et cohérent sans intervention humaine (Graefe, 2016). Dans le domaine du journalisme, ces logiciels ont pris les noms de robots-journalistes ou encore de robots-rédacteurs. On utilise également fréquemment le terme journalisme automatisé (Latar, 2018), qui a été d'ailleurs anticipé par Michael Frayn dans sa nouvelle satirique *The Tin Men* en 1965 (Frayn, 1965). Dès 2010, en réalisant cette ancienne utopie, les technologies de génération automatique de textes ont fait leur apparition en presse écrite (Latar, 2018). Ces logiciels, déjà bien ancrés aux États-Unis, ont petit à petit été introduits en Europe (Thurman, 2019). La France s'est révélée comme un acteur précurseur, mais d'autres pays européens ont désormais suivi le mouvement (Dierickx, 2019). Comme nous le détaillerons à travers les exemples de *SudPresse* et de *Mediafin*, la Belgique constitue à ce jour un nouvel acteur non négligeable dans le domaine du journalisme automatisé.

2. L'ascension des robots

Ray Kurzweil, ingénieur et futurologue au sein de l'éminente société américaine Google, a récemment annoncé⁶ que d'ici 2029, les ordinateurs seront non seulement capables de réaliser tout ce que les humains savent faire, mais qu'ils le feront mieux (Levy, 2013). Lors d'une conférence organisée récemment à Toronto⁷, Daniel Kahneman (2017), lauréat du prix Nobel d'économie en 2002, a validé la position Kurzweil. Il y a affirmé qu'il ne voyait aucune raison pour laquelle il resterait un jour une seule tâche que les robots ne seraient pas capables de réaliser. Par ailleurs, selon lui, une IA sera toujours plus efficace qu'un humain, car cette dernière ne sera jamais dérangée par ce qu'il qualifie de « bruit » (*noise*). Les humains, eux, sont sans cesse perturbés par ces interférences. D'après Kahneman, reconnaître l'existence de ce bruit a plusieurs implications pratiques :

One implication is obvious: you should replace humans by algorithms whenever possible. Even when the algorithms don't do very well, humans do so poorly and are so noisy that just by removing the noise, you can do better than people (Kahneman, 2017)⁸.

⁶ Au cours d'une interview menée par Steven Levy pour le magazine américain *Wired*. Cette interview faisait suite à un live *Google hangout*, animé par Ray Kurzweil, pour la sortie du film *After Earth* (Levy, 2013).

⁷ Lors d'une conférence sur l'IA tenue par l'association *Economics of AI* le 13 septembre 2017 à Toronto.

⁸ Une de ces implications est évidente : il faut remplacer les humains par des algorithmes dès que possible. Même lorsque les algorithmes ne sont pas très avancés, les humains sont tellement mauvais et dérangés par des interférences extérieures (du bruit), que, juste en retirant ce bruit, les algorithmes sont plus efficaces que les humains (notre traduction).

C'est précisément cet enjeu qui a animé ce travail de fin d'études. Appliqué au domaine du journalisme, cela signifierait-il que des robots d'IA seront, un jour, capables de remplacer entièrement les journalistes de presse écrite ? L'ère des journalistes humains est-elle bientôt révolue ? Plus encore, les robots feront-ils un jour un meilleur travail que leurs homologues humains ? Ou est-on face à une prévision digne d'un roman de science-fiction ? Loin de répondre de façon précise à ces questions périlleuses sur lesquelles les avis ne s'accordent pas, ce mémoire vise, à son échelle, à retracer un historique problématisé de l'introduction des générateurs automatiques de textes dans le domaine du journalisme de presse écrite, de sorte à mettre en lumière ses divers enjeux. Ainsi, nous tenterons de clarifier son utilisation, mais également d'aborder les problèmes engendrés par ces générateurs. Leur introduction dans le domaine du journalisme pose en effet des questions éthiques, déontologiques et légales qui seront discutées tout au long de ce mémoire.

3. Transformation, digitalisation et automatisation du monde des médias

Lors d'une interview réalisée dans le cadre de ce travail, Émilie Mascia, rédactrice en chef du magazine *ELLE à table Belgique*⁹, affirmait :

Aujourd'hui, faire un article web, ce n'est plus seulement copier-coller un texte qu'on avait fait pour le magazine papier et le mettre en ligne. [...] Il y a tout un travail qui est beaucoup plus poussé qu'avant. [...] C'est presque un nouveau métier. En tout cas, on voit que les journalistes qui sont là depuis 15 ans ont un peu plus de mal à s'adapter à cette nouvelle façon de travailler (entretien du 20 janvier 2020 avec Émilie Mascia).

Entre les textes alternatifs, l'adaptation des titres pour le web ou encore l'utilisation répétitive de mots-clés pour s'assurer du référencement optimal d'un article sur Google, le métier de journaliste de presse s'est effectivement métamorphosé. La démocratisation d'Internet, l'ascension de GAFAM¹⁰ et la diversification des plateformes numériques ont fait apparaître des facettes du métier qui ne faisaient jusqu'alors pas partie du quotidien des journalistes. De façon générale, durant ces dernières années, le monde des médias a subi une transformation rapide et radicale. La consommation des informations se fait de plus en plus en ligne, n'importe où et sur une multitude d'appareils différents (Einav, 2015). Parallèlement, les

⁹ Déclinaison culinaire du magazine *ELLE Belgique* publié aux éditions *Ventures* au nombre de deux tirages papier par an. Le site internet d'*ELLE à table Belgique*, quant à lui, est alimenté chaque jour.

¹⁰ Voir glossaire (p.5).

entreprises médiatiques se sont mises à produire toujours plus et à n'importe quelle heure de la journée sur leurs plateformes en ligne. Les stratégies SEO se sont développées et ont radicalement modifié la façon de concevoir un article (Mercier & Pignard-Cheynel, 2014). Avec de moins en moins de temps et plus de contraintes, les journalistes se plaignent souvent de manquer de temps pour produire un travail de qualité. Des interviews réalisées dans le cadre de ce travail, il est ressorti que rares sont les moments où ils ont l'occasion de s'investir longuement sur un sujet. Dans ce contexte, l'IA et les générateurs automatiques de textes peuvent apparaître comme une solution, facilitant le travail de recherche, soulageant les journalistes d'un travail répétitif et leur permettant ainsi de consacrer plus de temps aux articles qui leur tiennent à cœur, les robots-journalistes s'occupant d'une partie rébarbative du travail. Cette solution se montre bien évidemment controversée ; tous ne l'accueillent pas de manière positive et certains la considèrent comme une véritable menace pour le journalisme humain (Latar, 2018).

L'intégration des algorithmes en journalisme peut être retracée en trois étapes : le *computer assisted-journalism*, le datajournalisme et le robot-journalisme (Coddington, 2015 ; Latar, 2018). Les avancées technologiques ont toujours transformé le journalisme de manière considérable au cours du temps (Wu et al., 2019). *Computer assisted-journalism* renvoie ainsi au premier stade, qui correspond à l'introduction d'ordinateurs dans le processus de réunion des sources du journaliste, dès 1950. Dans les années 1980, l'utilisation d'ordinateurs dans les salles de rédaction s'est démocratisée et popularisée, jusqu'à devenir indispensable aujourd'hui. À ce stade, les ordinateurs étaient utilisés principalement pour réaliser des recherches, effectuer des statistiques et composer un article (Cox, 2000). Le datajournalisme, quant à lui, correspond à l'utilisation d'algorithmes pour trouver des informations cachées dans d'énormes *data silos*¹¹. La pratique s'est aujourd'hui démocratisée et est employée par la plupart des grandes entreprises médiatiques (Stone, 2014). Le robot-journalisme correspond au stade actuel de l'évolution. Les algorithmes sont aujourd'hui utilisés à toutes les étapes du travail journalistique, pour trouver des informations et pour les transformer en articles journalistiques sans intervention humaine. Les robots-journalistes peuvent être composés d'un simple programme informatique, d'une structure 3D et même posséder une apparence humaine (Latar, 2018). À côté de cela, les compagnies médiatiques utilisent de plus en plus l'IA à des fins marketing, en créant une information taillée sur mesure en

¹¹ Voir glossaire (p.4).

fonction d'un public cible (Targnion, 2019). L'utilisation de *Big Data* permet effectivement aux entreprises médiatiques d'obtenir des données telles que la location d'une personne, la façon dont elle utilise son téléphone, les sites et les applications qu'elle utilise, les moments où elle est connectée, les informations qu'elle poste sur Facebook, ou encore ce qu'elle aime et partage sur Instagram. Ainsi, un journal peut utiliser ces données pour produire des articles spécifiques sur base des préférences d'une personne, augmentant alors la pertinence de l'information pour le lecteur et s'assurant de sa fidélité (Ruskin, 2018).

Qu'on la considère comme une opportunité ou une menace, l'IA est en train de conduire à une modification de nombreux aspects du métier ; des procédures automatiques ont été introduites pour couvrir toutes les étapes du journalisme classique, de l'investigation à la distribution, en passant par la production de contenu. Au cœur de certaines entreprises médiatiques, les rôles des humains dans ces domaines ont déjà été entièrement ou partiellement remplacés par des processus automatiques et des robots (Latar, 2018). Dans le contexte de ce travail, nous nous intéresserons uniquement aux générateurs automatiques de textes, en laissant de côté les autres utilisations de l'IA, tels que les robots à téléprésence, l'utilisation des algorithmes à des fins marketing ou encore la réalisation automatisée de vidéos. Le nombre de cas dans lesquels l'IA est utilisée se révèle très vaste. Nous ne nous pencherons ainsi que sur une infime partie des capacités de l'IA en journalisme, l'automatisation de textes, et tenterons de la mettre en lumière.

4. Enjeux de l'utilisation de la génération automatique en presse écrite

Nous établirons un processus de mise sur agenda¹², comme défini par le politologue Patrick Hassenteufel (2010), de la question des générateurs automatiques de textes dans le domaine de la presse écrite. Nous participerons ainsi à la « mise en visibilité » du problème « par ses différents acteurs » (Hassenteufel, 2010, p.53). Bien qu'elle soit à ce stade encore relativement inexplorée, tout porte à croire que cette problématique continuera progressivement à s'imprégner dans l'agenda des problèmes publics (Latar, 2018). Effectivement, si la technologie de génération automatique de textes n'en est encore qu'aux prémices de son développement, le journalisme automatisé a déjà trouvé sa place au cœur

¹² Concept de sociologie politique que nous appliquons, dans le cadre de ce travail, aux sciences de la communication.

des rédactions et est très susceptible d'y perdurer (Graefe, 2016). Ses enjeux nous concernent tous, aussi bien en tant que citoyens, journalistes, que consommateurs et auront, d'une manière ou d'une autre, des conséquences sur les pratiques journalistiques futures au sein des entreprises médiatiques (Neveu). Le journalisme n'a eu de cesse de se renouveler au cours du temps (Cox, 2000 ; Mercier & Pignard-Cheynel, 2014) et la génération automatique de textes va, encore une fois, bousculer le quotidien des journalistes, les obligeant à reconsidérer leurs rôles et leurs aptitudes (Van Dalen 2012). Il est clair que ces nouvelles technologies vont également avoir d'importantes conséquences sur la manière dont nous consommons les médias (Ruskin, 2018). Par ailleurs, le journalisme automatisé prend place dans un contexte économique fragile pour la presse (Franklin, 2008). Si la question de la génération automatique de textes en journalisme n'est pas abordée de manière publique, la confiance des lecteurs, déjà fragile, pourrait encore s'amoinrir (Franklin, 2008 ; Neveu, 2010).

Les enjeux se révèlent donc ici multiples. Ils concernent à la fois les pratiques journalistiques, la condition des journalistes, l'organisation des rédactions et la réception des lecteurs. Plus largement, ils concernent également la place de l'IA dans le domaine du journalisme. À terme, le métier est-il menacé par les technologies d'automatisation de textes ? S'agit-il au contraire d'une opportunité, à la fois pour les entreprises médiatiques et les journalistes ? Comment les journalistes vont-ils cohabiter avec les robots-journalistes ? Par ailleurs, les générateurs automatiques sont-ils capables d'appliquer les principes déontologiques et éthiques ? Peuvent-ils être créatifs ? Ces divers enjeux seront au centre de ce mémoire.

5. Trois temps pour poser un regard sur l'IA en presse écrite

Ce travail est divisé en trois sections distinctes, qui permettront d'aborder les questions essentielles relatives aux technologies de génération automatique de textes en presse écrite. Tout d'abord, après avoir brièvement détaillé le fonctionnement de ces logiciels, le premier chapitre retracera un historique problématisé de leur développement selon les différentes parties du globe. Nous nous attarderons effectivement sur leur utilisation aux États-Unis, en Chine, au Japon, en France et dernièrement en Belgique. Nous avons fait le choix de détailler plus longuement les façons dont la génération automatique de textes est, ou sera bientôt, employée dans les rédactions belges en se basant sur des entretiens réalisés avec les chefs de projets des rédactions de *SudPresse* et de *Mediafin*. Nous aborderons ainsi les raisons pour lesquelles ces technologies ont fait leur entrée dans la profession et les conditions de leur

utilisation. De manière générale, cette première section offrira une base historique et théorique permettant de cerner au mieux les enjeux explorés dans le deuxième chapitre.

La deuxième étape du travail aura pour objectif principal de dresser les différentes conséquences induites par la génération automatique de textes en presse écrite. Dans un premier temps, nous aborderons les questions éthiques qu'elle engendre, notamment le sujet de la transparence et les règles de déontologie journalistique. Ensuite, nous nous pencherons sur les enjeux juridiques de son utilisation, un cadre légal n'ayant, à ce jour, pas encore été établi concernant l'emploi de l'IA. Nous analyserons également les différentes menaces que la génération automatique de textes pourrait poser et les contrebalancerons avec les opportunités qu'offre cette dernière, non seulement aux entreprises médiatiques, mais aussi aux journalistes. Nous adresserons ensuite la question de la survie du métier de journaliste, de nombreuses craintes ayant été émises à ce sujet. Pour ce faire, nous aborderons les limites des algorithmes de génération automatique de textes et tenterons de déterminer si ces dernières ouvrent de nouveaux horizons aux journalistes humains. Cette deuxième section sera illustrée par des propos recueillis auprès de différents acteurs de l'industrie du journalisme.

Enfin, le troisième chapitre consistera en une étude de contenu d'articles générés automatiquement. Dans cette partie, nous avons, à nouveau, choisi de nous intéresser aux rédactions belges en analysant des articles générés automatiquement par *SudPresse*. Les articles choisis consistent en de courts résumés de matches de football amateur, qui sont généralement publiés au nombre de 6 000 par semaine sur le site internet de *SudPresse* (entretien du 30/03/2020 avec André Thiel, Directeur *Process & Development* chez *SudPresse*). Grâce à l'étude de contenu de trente de ces articles, nous aspirons à illustrer clairement la forme que les textes générés automatiquement peuvent prendre, leur structure, ainsi que leurs rouages intrinsèques. Cette section permettra également de souligner ce que nous avons appelé la « plus-value » journalistique, en comparant des articles automatisés à des articles rédigés traditionnellement par des journalistes humains. Ainsi, à travers ces trois chapitres, nous espérons couvrir le sujet de manière complète et percutante. De manière plus générale, nous souhaitons également souligner les enjeux, aussi bien technologiques que sociaux, de l'utilisation de l'IA au sein de notre société.

6. Méthodologie

Ce mémoire est le fruit d'une méthode de recherche classique, impliquant lectures académiques, articles de presse, conférences et sources provenant d'Internet. Les recherches ont porté sur plusieurs combinaisons de mots-clés à travers différents opérateurs, tels que Google Scholar, Google et le moteur de recherche de la bibliothèque en ligne de l'Université de Liège. L'étude de contenu du troisième chapitre, elle, a été inspirée du livre *Analyser le discours de presse*, de Roselyne Ringoot (2014) et du tableau d'analyse que cette dernière propose en fin d'ouvrage. Ce travail a également été enrichi par des entretiens poursuivis auprès de journalistes ou d'acteurs du monde de la presse, majoritairement réalisés par téléphone. Ces interviews ne se veulent aucunement représentatives des différents avis émis sur l'introduction de l'IA en presse écrite. Ils ne prétendent pas non plus représenter un échantillonnage équilibré des avis exprimés sur le sujet. Ils sont ainsi utilisés dans l'unique but d'illustration, ou, dans certains cas, en tant que compléments de sources.

Ont été interviewés dans le cadre de ce mémoire : Marie Guérin (rédactrice en chef d'*ELLE Belgique*), Émilie Mascia (rédactrice en chef d'*ELLE à table Belgique*), Nadia Chater (journaliste aux *Éditions L'Avenir* et coordinatrice de *Deuzio*), Ricardo Gutiérrez (secrétaire général de la Fédération Européenne des Journalistes et membre du Conseil de Déontologie Journalistique), André Thiel (directeur *Process & Development* chez *SudPresse*), Nicola Becquet (responsable de la stratégie numérique de *L'Écho*), Julien Covolo (journaliste web à la RTBF), Xavier Ducarme (rédacteur en chef adjoint de *La Libre*), David Coppi (journaliste politique au *Soir*) et Claude de Loupy (CEO de la société de génération automatique de textes Syllabs). Toutes les retranscriptions des entretiens se trouvent dans les annexes de ce travail.

7. Genèse et limites du mémoire

Avant d'entrer dans le cœur du sujet dans les pages qui suivent, nous souhaitons évoquer l'une des réalités à laquelle notre recherche a été confrontée : la confusion et l'étonnement face au sujet de recherche choisi. Notre entourage a effectivement souvent eu du mal à établir clairement le lien entre IA et journalisme. La plupart des journalistes interrogés dans le cadre de ce travail se sont également trouvés décontenancés face à l'évocation de la génération automatique de textes en presse écrite. Nombre d'entre eux ignoraient ces techniques et

étaient encore moins informés du fait qu'elles sont utilisées dans le domaine du journalisme. L'une des journalistes interviewée a même douté de leur existence : « Je ne savais pas que cette technologie existait. Ça existe vraiment ? » (entretien du 30 mars 2020 avec Nadia Chater, journaliste pour *L'Avenir* et coordinatrice de *Deuzio*). Ce type de remarque ne s'est pas trouvé isolé. La génération automatique de textes est pourtant bel et bien utilisée en journalisme depuis une dizaine d'années, et plus de cinq ans en Europe. S'ils en connaissaient les principes, les journalistes interviewés ne savaient souvent pas comment cette technologie fonctionnait réellement. Cette méconnaissance du domaine, tant au niveau de nos proches, que des journalistes, nous ont poussé à continuer nos recherches et à confirmer la pertinence du sujet. Dans un monde où l'IA et les générateurs automatiques de textes se trouvent en pleine évolution, il nous a semblé important de s'y pencher de plus près, d'en comprendre les mécanismes et d'adresser les opportunités et les menaces qu'ils posent au journalisme de presse écrite. Ainsi, nous espérons, qu'à son échelle, ce mémoire permettra de mettre en lumière un domaine encore relativement méconnu.

Les limites de ce travail sont premièrement de nature quantitative. Nous avons fait le choix de limiter l'historique du premier chapitre aux innovations américaines, asiatiques et française les plus marquantes, ainsi qu'à la Belgique. Nous avons donc délibérément choisi de ne pas aborder en détail les productions du continent océanique ou d'autres pays européens tels que l'Allemagne ou les pays nordiques, qui se trouvent aussi être des acteurs prépondérants. Enfin, ce mémoire n'abordera volontairement que très peu la réception des articles générés automatiquement, qui a déjà été adressée dans plusieurs études récentes et qui semble converger vers les mêmes conclusions (Graefe et al., 2018)¹³. Nous avons ainsi pris le parti de traiter la génération automatique de textes du point de vue de la production, des journalistes et des entreprises médiatiques.

¹³ Les recherches concernant la réception des lecteurs sont notamment discutées dans les études suivantes : Clerwall, C. (2014). Enter the Robot Journalist. *Journalism Practice*, 8(5), pp.519-531.
Graefe, A., Haim, M., Haarmann, B., & Brosius, H. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, 19(5), pp.595-610.
Jung, J., Song, H., Kim, Y., Im, H., & Oh, S. (2017). Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. *Computers in Human Behavior*, 71(C), pp.291–298 .

CHAPITRE 1

Historique problématisé

1.1. Avant-propos : âge de la singularité, IA faible et IA forte

En 1993, l'écrivain de science-fiction américain Vernor Vinge prédisait que nous aurions les moyens technologiques de créer une IA dépassant les capacités humaines en moins de trente ans à venir. Selon lui, cette période signerait la fin de l'ère humaine (Godwin, 2007). Si en 2020, sa prédiction ne s'est pas encore vérifiée, l'idée de Vinge a depuis lors été théorisée. Ray Kurzweil, futuriste éminent au sein de la société américaine Google, définit la « singularité » comme le moment où les hommes et les machines convergeront. Il s'agit selon lui d'un événement singulier, qui n'aura lieu qu'une seule fois dans l'histoire humaine (Shanahan, 2015). L'âge de la singularité implique effectivement que des développements technologiques exponentiels apportent tant de changements que la société, au sens où nous l'entendons aujourd'hui, prend fin. Les valeurs humaines les plus basiques en seraient bousculées. Nos compréhensions de l'humanité, de l'individu, de la conscience, seraient également remises en question par ces nouvelles circonstances (Shanahan, 2015).

Si l'âge de singularité peut encore paraître lointain, il est pertinent de souligner que les développements dans le domaine de l'IA se concrétisent à une vitesse incroyable (Kahneman, 2017). Certaines évolutions dépassent d'ailleurs déjà notre compréhension, l'homme n'étant pas toujours en mesure d'expliquer la manière dont les algorithmes évoluent, s'autocorrigent et pensent. En 2017, des chercheurs appartenant à la société Facebook programmaient deux *chatbots*¹⁴ de façon à ce qu'ils s'entraînent à la négociation commerciale. Après les avoir laissés discuter entre eux quelques heures, les *chatbots* ont créé leur propre langage, leur permettant de communiquer, d'échanger et de négocier de manière optimale. De façon paradoxale, les concepteurs des *chatbots* se sont retrouvés dans l'incapacité de comprendre ce langage développé par les logiciels d'IA, qu'ils avaient pourtant eux-mêmes programmés. Grâce aux algorithmes de *Machine Learning*¹⁵, les deux *chatbots* ont effectivement acquis des compétences propres, qui nous sont incompréhensibles. De plus, alors qu'ils n'étaient

¹⁴Agents capables de simuler un dialogue avec des utilisateurs humains par échange textuel ou vocal. Ces outils se sont notamment démocratisés sur les sites internet commerçants et sur Facebook durant ces dernières années (Zarouali et al., 2018).

¹⁵ Voir glossaire (p.5).

pas programmés pour, les robots ont développé des capacités avancées de ruse et de stratégie (Moreira, 2017). Comme cet exemple le démontre, les développeurs d'IA sont aujourd'hui incapables de déterminer le raisonnement de prise de décision de leurs algorithmes. Nous sommes effectivement encore loin de comprendre les mécanismes complexes du *Deep Learning*¹⁶ et, de manière plus générale, les capacités potentielles de l'IA (Latar, 2018).

Avant d'atteindre l'âge de la singularité, des progrès conséquents sont encore à réaliser. Malgré tous les avancements dans le domaine, incluant l'introduction des algorithmes intelligents¹⁷, des réseaux neuronaux¹⁸, des méthodes de *Deep* et de *Machine Learning*, les algorithmes les plus sophistiqués à ce jour sont de type faible¹⁹, incapables de réaliser plusieurs tâches simultanément. L'âge de la singularité sera précédé par le développement d'une IA forte²⁰, qui consistera en un être artificiel doté d'une conscience, d'un esprit, ou encore d'une sensibilité. Ce stade n'a encore jamais été atteint²¹. Selon les futurologues, l'étape suivante consistera en le développement d'une « super IA », dont l'intelligence et la créativité dépasseront celles des humains (Dickson, 2017). Nous ne sommes pour ainsi dire qu'à l'aube du développement potentiel de l'IA (Kahneman, 2017). Même si nous n'avons à ce jour développé que des logiciels de type faible, ces derniers ont cependant déjà envahi tous les aspects de la communication et des organisations humaines. Parmi ces systèmes d'IA faible, on retrouve notamment les algorithmes de génération automatique de textes, qui ont pris une place singulière au sein du versant médiatique durant ces dernières années (Latar, 2018). Ce sont ces algorithmes spécifiques qui nous intéresseront tout particulièrement tout au long de ce travail.

¹⁶ Voir glossaire (p.4).

¹⁷ Voir glossaire (p.3).

¹⁸ Voir glossaire (p.6).

¹⁹ Voir glossaire (p.3).

²⁰ Voir glossaire (p.3).

²¹ En 2014, IBM a commencé à réaliser des recherches dans le but de se rapprocher le plus possible de développement d'une IA forte. La société a développé une nouvelle puce d'ordinateur neuro-synaptique destinée à aller au-delà des limitations des procédés d'IA faible, en intégrant des capacités de la partie gauche et droite du cerveau humain sur le principe amélioré des réseaux neuronaux. S'il se réalise, le cerveau IBM pourrait être considéré comme une aide future à la sécurité publique, la surveillance médicale, aux transports ou encore à l'écriture d'articles journalistiques (Latar, 2018).

1.2. Développement des générateurs automatiques de textes

1.2.1 Bref historique

La génération automatique de textes s'est développée au lendemain de la Seconde Guerre mondiale. En 1948, Claude Shannon, ingénieur en génie électrique, publiait un article intitulé *Théorie mathématique de la communication*, qui ouvrait la discussion²². Inspiré par l'article de Shannon, le mathématicien de la Fondation Rockefeller, Warren Weaver, a entrepris de réunir les fonds nécessaires pour développer les prémices de la génération automatique de textes. Il finira cependant par abandonner son projet, notamment suite au constat de l'insuffisance des connaissances dans les domaines de la linguistique et de l'informatique. Son postulat, néanmoins, ouvrira la voie à de nombreuses autres recherches, dont les travaux du linguiste américain Noam Chomsky, qui sont à l'origine d'avancées significatives dans le domaine à la fin des années 1960 (Dierickx, 2015). Ces années verront ainsi les développements clés de la génération automatique de textes, mais ce n'est qu'à la fin des années 1970 que les premiers systèmes seront réellement mis au point. Les tentatives antérieures sont plutôt considérées comme des expériences théoriques (Danlos, 1985). Enfin, selon Claude Ponton (1997), c'est au cours des années 1980 que les générateurs automatiques de textes, à proprement parler, feront leur apparition et produiront des phrases intelligibles et compréhensibles pour les lecteurs. C'est également à ce moment que les logiciels élargiront leurs compétences au-delà de l'Anglais (Ponton, 1997). Dans le courant des années 1990, plusieurs réalisations ayant été poursuivies avec succès, le domaine suscitera un véritable engouement et ne cessera de se développer au fil de l'amélioration des techniques (Dierickx, 2015).

1.2.2 Fonctionnement général

La génération automatique de textes est ainsi apparue comme un nouveau domaine de recherche dirigé par des logiciels d'IA qui utilisent des connaissances importantes en linguistique et en étude du langage naturel. En narration automatique, les algorithmes sont utilisés pour convertir automatiquement, sans intervention humaine et en une fraction de

²² Shannon y analyse les schémas généraux des modèles de la communication. Concernant la génération automatique de textes, son approche mathématique se base sur le processus stochastique, aussi appelé processus aléatoire, au sein d'enchaînements de mots fondés à partir de chaînes de Markov. Il y étudie les probabilités d'enchaînements de mots qui forment des phrases grammaticalement correctes en Anglais (Shannon, 1948).

seconde, des données provenant de *data silos* en des textes lisibles (Latar, 2018). Cette technologie repose notamment sur le *Big Data* et le *Natural Language Generation* (NLG)²³ (Graefe, 2016 ; Reiter & Dale, 2000). Par ailleurs, il est pertinent de mentionner que la génération automatique de textes se rapproche du domaine de *Natural Language Processing* (NLP)²⁴, dont le but est de permettre aux systèmes d'IA de traiter et de comprendre le langage humain de façon intelligente. Ces deux domaines de recherche, qui paraissent opposés, sont en réalité très similaires. NLP et NLG partagent les mêmes fondements théoriques et sont notamment souvent utilisés conjointement au sein de divers programmes (Dale & Reiter, 2000).

Les générateurs automatiques de textes fonctionnent de la manière suivante : le logiciel reçoit en entrée (input) des données qu'il analyse et produit en sortie (output) un texte lisible et cohérent sans intervention humaine. Les textes générés doivent en ressortir cohérents, pertinents et grammaticalement corrects (Danlos, 1985). Selon Danlos, (1985), ces logiciels présentent de nombreux avantages, tels que la flexibilité et la personnalisation. À partir des mêmes données, un logiciel peut effectivement générer plusieurs phrases différentes, et notamment en plusieurs langues s'il a été programmé pour. Les logiciels ont ainsi l'avantage de pouvoir être ajustés selon les envies et les besoins de l'utilisateur. D'après Dale et Reiter (2000), le coût représente un autre avantage important à prendre en compte. Même si le développement ou l'acquisition des logiciels représente un investissement important, l'utilisation des générateurs automatiques de textes resterait plus économique que d'employer une personne pour écrire manuellement. De plus, ces systèmes présentent l'avantage de produire des textes très rapidement, en quelques secondes seulement et sans erreur d'inattention (Dale & Reiter, 2000).

1.2.3 Une offre qui s'est diversifiée

Il existe aujourd'hui treize sociétés à travers le monde qui proposent des solutions automatisées de création de textes. Deux sont basées aux États-Unis (Automated Insights et Narrative Science) et sont considérées comme les pionnières du mouvement (Latar, 2018). Cinq sont basées en Allemagne (2txt NLG, AX Semantics, Textomatic, Text-On, Retresco), deux en France (Syllabs, LabSense), une en Espagne (Narrativa), une au Royaume-Uni

²³ Voir glossaire (p.5).

²⁴ Voir glossaire (p.5).

(Arria), une en Russie (Yandex) et une en Chine (Tencent). Plusieurs d'entre elles sont capables de générer du texte dans plusieurs langues, comme la compagnie allemande AX Semantics, qui offre ses services d'automatisation en douze langues différentes (Graefe, 2016 ; Dierickx, 2019). Il est important de mentionner qu'aucune de ces entreprises ne se revendique en tant qu'organisation journalistique, leurs logiciels n'étant pas destinés spécifiquement et uniquement au journalisme automatisé (Graefe, 2016).

1.3. Développement du journalisme automatisé

1.3.1. Une utilisation de la génération automatique de textes parmi d'autres

À ce jour, la génération automatique de textes est utilisée dans de multiples domaines et à de nombreuses fins différentes. Ces techniques peuvent être appliquées dans n'importe quel cadre tant que des données structurées sont disponibles (Graefe, 2016). Ainsi, les logiciels sont notamment utilisés pour générer des rapports d'entreprises, des contenus de sites internet, des textes destinés aux réseaux sociaux, des descriptions de lieux dans le secteur de l'immobilier ou encore des notices d'utilisation ou des rapports de patients hospitalisés dans le milieu médical. L'automatisation des articles journalistiques ne constitue dès lors qu'une des multiples possibilités de la génération automatique de textes parmi beaucoup d'autres (Dierickx, 2015).

Selon Amir Ruskin (2018), l'invention de *Big Data* constitue une véritable aubaine pour les journalistes et les médias, qui ont la possibilité d'exploiter le pouvoir des données pour ajouter de la valeur au journalisme (Ruskin, 2018). Alors que les ordinateurs ont longtemps assisté les journalistes dans les différentes phases du processus de production, telles que la récolte, l'organisation de données ou l'analyse de *data silos*, l'écriture était jusqu'alors restée du domaine exclusif des journalistes (Graefe, 2016). La génération automatique de textes a bousculé ce mode de fonctionnement en introduisant le journalisme automatisé. Ce dernier est défini comme le fait d'utiliser des algorithmes pour générer automatiquement des articles sans intervention humaine (Graefe, 2016). Ce nouveau domaine du journalisme robotique est basé sur deux piliers : le logiciel qui extrait automatiquement des nouvelles connaissances de *data silos* et des algorithmes qui convertissent automatiquement, et sans intervention humaine, ces données en connaissances lisibles (Latar, 2018). Les entreprises médiatiques qui souhaitent utiliser ces logiciels ont le choix entre les différentes entreprises extérieures

citées plus haut, mais elles peuvent également choisir de développer leurs propres algorithmes en interne. Des organisations importantes comme *The New York Times* et le *Los Angeles Times* ont créé des départements qui, en leur sein, font la transition vers le digital et conçoivent leur propres logiciels. D'autres, comme *Reuters*, s'associent à des tierces parties pour combler ce vide dans l'évolution très rapide du marché du journalisme. Il s'agit là de deux exemples illustrant la manière dont les entreprises médiatiques travaillent avec les technologies de *Big Data* aujourd'hui. Certaines recrutent directement leur propres équipes de *data scientists*²⁵, alors que d'autres ont choisi de faire appel à la sous-traitance (Ruskin, 2018). Ces différents cas de figure seront abordés dans les pages suivantes.

1.3.2. Premiers pas du journalisme automatisé

La première expérience de génération automatique de textes dans le domaine du journalisme a été menée par la chercheuse française en linguistique et informatique Laurence Danlos. Dès 1985, cette dernière a créé un système automatique permettant de générer des articles succincts traitant des attentats sur un ton journalistique. Ces courts textes, créés sur base de données envoyées via formulaires, ressemblaient beaucoup aux articles des dépêches d'agence (Danlos, 2010). Cette première tentative de développement du journalisme automatisé, lancée en France, s'est matérialisée vers la fin des années 2000, lorsque des algorithmes d'IA ont été développés à la *Northwestern University's computer science and communication labs*, avec pour objectif d'aider les journalistes humains à rédiger des articles (Latar, 2018). Un changement majeur a également eu lieu lorsque Narrative Science et Automated Science, les deux pionniers du journalisme automatisé aux États-Unis, ont commencé à commercialiser leurs services auprès des entreprises médiatiques (Dierickx, 2015).

1.3.3. Fonctionnement général

La *figure 1* illustre, de manière basique, comment fonctionne la génération automatique de textes en journalisme. En premier lieu, le logiciel d'IA collecte les données disponibles. Dans le cas d'un article résumant un match de sport par exemple, il pourrait s'agir des résultats du match, des déplacements en temps réel des joueurs ou encore des records historiques des différentes équipes (Graefe, 2016). Ensuite, les algorithmes utilisent des méthodes

²⁵ Voir glossaire (p.4).

statistiques pour identifier les informations importantes au cœur de ces différentes données (Thurman, 2019). Toujours dans le cas d'un match de sport, cela pourrait s'avérer être la performance extraordinaire d'un joueur, un événement inhabituel ou encore le moment décisif du résultat du match. En troisième lieu, le logiciel classe les informations identifiées par importance, de façon à ne conserver que celles qui intéresseront réellement les lecteurs. Par la suite, les algorithmes génèrent un texte sur base des éléments choisis en suivant les règles par lesquelles ils ont été programmés. Enfin, l'article pourra être téléchargé sur le système de gestion de contenu de l'éditeur. Il sera alors soumis à relecture ou publié automatiquement sans ajustements préalables, le choix étant à réaliser lors de la programmation (Dörr, 2016 ; Graefe, 2016). Le journalisme automatisé permet ainsi de publier des articles très rapidement, en quelques secondes, dès que les données sont disponibles (Graefe, 2016).



Figure 1. Le processus de génération automatique d'articles de presse (Graefe, 2016, p.18).

Durant le processus de production, le logiciel d'IA repose sur une série de règles spécifiques, prédéfinies dans le cas du sujet couvert. Ces règles sont souvent créées à la fois par des *data scientists*²⁶, des journalistes et des linguistes spécialisés en informatique. Les linguistes, en se basant sur des échantillons de textes du domaine étudié, doivent identifier la logique sémantique fondamentale de ceux-ci pour les traduire en un système capable de construire des phrases. S'il n'existe aucun échantillon de texte dans le domaine choisi, des journalistes peuvent également rédiger des modèles d'articles-types et les ajuster au style éditorial du

²⁶ Voir glossaire (p.4).

journal ou du site internet destiné à accueillir les articles. Des experts du domaine abordé par l'article (sport, finance, etc.) doivent également participer à l'élaboration des règles et définir à l'avance quels éléments sont d'intérêt journalistique. En reposant sur ces éléments, le logiciel sera capable de déterminer quelles informations recueilleront l'intérêt des lecteurs (Graefe, 2016).

1.3.4. Un travail de collaboration

Il est important de souligner le travail de collaboration que les *data scientists* réalisent lorsqu'ils instaurent des logiciels d'automatisation de textes au cœur d'entreprises médiatiques. Les algorithmes n'ayant pas été créés uniquement pour une utilisation dans le domaine de la presse écrite, ils nécessitent d'être ajustés et modifiés. Les entreprises de génération automatique de textes doivent effectivement adapter leurs logiciels à une utilisation dans le cadre spécifique de la production journalistique. Il est également nécessaire de les ajuster aux normes de l'institution médiatique désirant bénéficier de ses services. Spécifiquement, au cours de ce travail d'adaptation, les ingénieurs doivent prendre en compte les règles de leur propre terrain de connaissances, mais aussi les règles intrinsèques au journalisme (Wu et al., 2019). Nous pouvons envisager cette collaboration à l'appui de la théorie des champs du sociologue français Pierre Bourdieu (1996).

La théorie bourdieusienne des champs définit la société et l'ensemble des actions humaines comme une imbrication de champs. Ces derniers peuvent notamment être culturels, économiques, sportifs, artistiques ou encore religieux et sont organisés selon des logiques qui leur sont propres, qui dépendent des propriétés et des enjeux spécifiques à chaque champ. Selon Bourdieu, ces champs sont semi-autonomes et se trouvent donc soumis à des influences externes, les différents champs ayant la capacité de s'influencer les uns les autres. Les agents et les personnes physiques appartenant à un champ particulier ont également la capacité d'agir et de réagir pour conserver ou transformer un champ (Bourdieu, 1996).

Appliqué à notre domaine d'étude, le journalisme peut être considéré comme un champ particulier régi par un ensemble de règles qui influencent la façon dont les journalistes se comportent. Ces règles incluent notamment des principes éthiques, des règles de déontologie, des normes et des standards professionnels. Les actes des journalistes, influencés par leur expérience personnelle et professionnelle, peuvent notamment bouleverser et modifier la

logique du champ global. On peut également considérer que le champ du journalisme est influencé par différents autres champs dominants, tels que les champs politiques et économiques, par exemple²⁷(Wu et al., 2019). Dans le cas de notre étude, les entreprises de génération automatique de textes peuvent être considérées comme des nouveaux acteurs du champ journalistique, ou encore comme des agents qui font pression sur ce champ de par leurs offres. La sphère technologique et les *data scientists* (agents extérieurs) ont ainsi eu une influence sur la logique de fonctionnement du champ journalistique.

Toujours selon la même théorie, Bourdieu (1996) explique qu'au sein d'un champ particulier, les nouveaux acteurs peuvent être accueillis avec résistance, surtout s'ils cherchent à modifier entièrement le *statu quo* du champ en question. Cette théorie est vérifiée dans le cadre de l'introduction du journalisme automatisé, qui n'a effectivement pas été bien accueilli par certains journalistes, craintifs par rapport à ces nouvelles technologies. Ces derniers avancent notamment l'argument que les robots finiront par s'approprier leur travail (Graefe, 2016). Cet aspect sera abordé dans le deuxième chapitre de ce mémoire. Par ailleurs, Bourdieu (1996) explique qu'en pénétrant dans un champ nouveau, les acteurs extérieurs doivent s'adapter et jouer avec les règles internes de ce champ. C'est également ce qui a pu être observé dans le domaine du journalisme automatisé, où les entreprises de génération de textes et les ingénieurs ont dû ajuster leurs logiciels aux contraintes et aux règles internes du journalisme. Inversement, les normes du champ du journalisme ont inévitablement dû être adaptées aux règles associées au champ technologique, la vie des rédactions étant incontestablement modifiée par l'introduction des technologies d'automatisation en leur sein (Wu et al., 2019).

1.3.5. Une tendance grandissante

Par la pratique intensive, parfois risquée et coûteuse du journalisme d'investigation, les journalistes cherchent à révéler des faits nouveaux et des tendances sociales. Avec leur talent narratif, leur expérience, leur créativité, leur intuition et leurs valeurs, ils convertissent des informations en articles journalistiques pour leur audience. Les robots journalistes constituent une compétition rude à cette pratique traditionnelle du journalisme. En plus des

²⁷ Ces influences ont lieu notamment lorsque les annonceurs modifient le contenu d'un journal (champ économique) ou dans le cadre d'un contrôle du gouvernement sur le contenu éditorial (champ politique) (Benson, 2006).

économies potentielles en coût de main-d'œuvre, les robots journalistes omettent rarement des informations essentielles, ne sont jamais fatigués et, s'ils sont programmés objectivement, n'ont pas d'a priori. Ces aspects concrets ont conquis plusieurs entreprises de presse. L'introduction de l'IA dans le domaine du journalisme est considérée comme une tendance grandissante (Latar, 2018). Lorsque l'on demandait à Arthur Sulzberger, éditeur du *New York Times*, ce qu'il aurait désiré faire différemment au cours de la transition digitale, ce dernier répondait spontanément « Hire more engineers »²⁸ (Doctor, 2013, paragraphe 1). Les technologies d'IA ont ainsi rapidement intégré les rédactions journalistiques et l'utilisation de logiciels créés par des ingénieurs dans le processus de rédaction constitue désormais, pour certaines, un élément essentiel à leur survie (Latar, 2018).

Aujourd'hui, le journalisme automatisé est principalement utilisé pour générer des articles routiniers et répétitifs, lorsque des données structurées et exactes sont disponibles. Les robots journalistes ont débuté par rédiger des articles dans la sphère du sport, pour après se diriger vers les sphères financières. Désormais, ils sont également employés dans d'autres domaines plus sophistiqués, la seule condition étant l'existence et l'accessibilité à des données structurées (Graefe, 2016 ; Latar, 2018). Les algorithmes sont ainsi utilisés par de nombreux organes de presse, dont l'agence de presse généraliste mondiale *Associated Press*, qui génère automatiquement des articles sur les revenus des entreprises (Finley, 2015). Par ailleurs, plusieurs compagnies ont également développé des algorithmes qui ajustent les tons et les structures narratives de leurs articles en fonction de profils d'audience prédéfinis (Latar, 2018). Yahoo, par exemple, crée des articles de *fantasy sport*²⁹ adaptés à ses lecteurs (Finley, 2015).

Il a été estimé que dans les cinq à dix prochaines années, la majorité des articles journalistiques seront écrits par des robots (Latar, 2018). De plus en plus d'entreprises se dirigent effectivement vers l'automatisation d'articles. Cependant, Daewon Kim et Seongscheol Kim (2017) soulignent que, proportionnellement, ces entreprises médiatiques restent encore minoritaires. Même si la tendance se développe, peu d'entreprises utilisent actuellement les services de sociétés spécialisées dans la génération automatique de textes.

²⁸ Engager plus d'ingénieurs (notre traduction).

²⁹ Le « fantasy sport » désigne un jeu virtuel de simulation où les participants endossent le rôle d'actionnaires d'équipes sportives, constituées de joueurs virtuels ou réels. Les joueurs perdent et gagnent des points en fonction de la pertinence de leurs choix stratégiques (Beliën, Goossens, & Van Reeth, 2017).

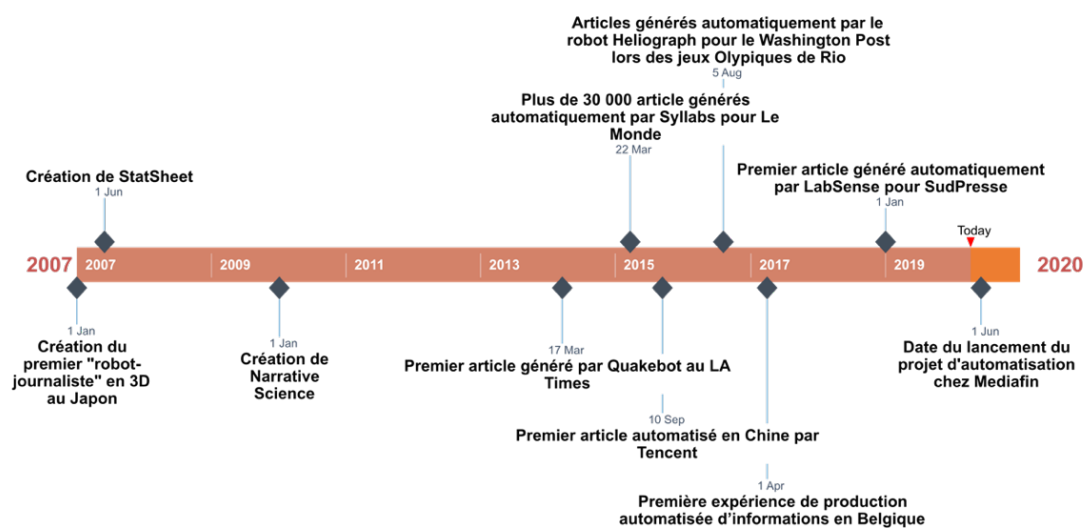
Par ailleurs, comme expliqué précédemment, la plupart des logiciels disponibles sont limités à la couverture de sujets routiniers, tels que le sport ou la finance, pour lesquels des données structurées et fiables sont disponibles (Kim & Kim, 2017). On peut donc considérer que le domaine du journalisme automatisé se trouve toujours en phase expérimentale, ou dans une phase précoce d'expansion de marché (Dörr, 2016). Cette phase pourrait néanmoins rapidement évoluer au vu des progrès constants réalisés dans les domaines de l'IA, de l'informatique, de l'analytique de *Big Data*, du NLP et du NLG (Stone, 2014 ; Graefe, 2016).

Un autre élément permettant d'affirmer l'expansion très probable du phénomène d'automatisation consiste en l'augmentation perpétuelle du nombre de données stockées. La quantité de *data silos*, qui sont le principal moteur du journalisme automatisé, ne cesse effectivement d'augmenter (Stone, 2014 ; Graefe, 2016). Ces données sont fournies notamment par des organisations, des détecteurs ou simplement par les utilisateurs du World Wide Web. Le monde est de plus en plus équipé de détecteurs en tous genres, qui génèrent et collectent automatiquement des données. À ce jour, des détecteurs traquent constamment la température de l'environnement, l'activité systémique ou encore le niveau de pollution de l'air. On utilise également constamment plus de détecteurs pour fournir des données sur des événements, tels que le nombre de personnes présentes au cours de ces derniers. Lors des matches de football américains, par exemple, la National Football League emploie des détecteurs pour connaître la position en temps réel de chaque joueur sur le terrain, leur vitesse, la distance qu'ils ont parcourue ou encore leur taux d'accélération. Par ailleurs, les utilisateurs des réseaux sociaux génèrent un nombre, en constante augmentation, de données en tous genres. Enfin, de plus en plus de pays tentent de rendre leur gouvernement plus transparent en lançant de nouvelles initiatives d'*open data*, de façon à rendre les données gouvernementales plus accessibles. Ce nombre exponentiel de nouvelles données favorise la potentielle démocratisation du journalisme automatisé, qui a un besoin structurel de données pour subsister (Graefe, 2016 ; Flew et al., 2011).

Aussi, en lumière de la situation précaire actuelle de l'industrie du journalisme de presse écrite et de la demande grandissante de contenu, il est pertinent de mentionner que les entreprises médiatiques sont constamment à la recherche de nouvelles manières de générer des revenus. L'automatisation d'articles routiniers pourrait notamment les aider à remonter la pente, tout en produisant toujours plus d'articles. Compte tenu de la capacité des robots-journalistes à produire du contenu en très grande quantité en quelques minutes à peine, le

journalisme automatisé apparaît, selon plusieurs chercheurs, comme une stratégie permettant aux entreprises médiatiques de diminuer leurs coûts de production et d'augmenter leurs bénéfices (Cohen, 2015).

Au vu de ces remarques, il n'est pas étonnant que les partisans du journalisme automatisé s'attendent à ce que le domaine se démocratise rapidement. En 2012, Kristian Hammond, cofondateur de Narrative Science, prédisait que dans les dix années à venir, plus de 90 % des articles seraient générés automatiquement, ce qui ne s'est pas réalisé (Levy, 2012). Saim Alkan, Directeur Général de la société allemande de génération de textes AX Semantics, estime qu'aujourd'hui déjà, les algorithmes seraient capables de produire la moitié du contenu d'un journal quotidien (Graefe, 2016). Ces points de vue sont bien évidemment discutables, d'autant qu'ils sont défendus par des personnalités qui ont des intérêts financiers dans le succès du journalisme automatisé. Cependant, l'automatisation étant utilisée par plusieurs entreprises influentes, telles qu'*Associated Press* et *Forbes*, la probabilité que d'autres suivent le pas rapidement se révèle très importante (Graefe, 2016). Selon Konstantin Dörr (2016), il existe des signes qui montrent que de plus en plus d'entreprises médiatiques se dirigent vers le journalisme automatisé. Il explique notamment que la plupart des sociétés de génération automatique de textes se trouvent en négociation constante avec des entreprises médiatiques intéressées par leurs produits. De plus, beaucoup d'entre elles refusent souvent de dévoiler l'identité de leurs clients du monde des médias. Le journalisme automatisé pourrait donc déjà s'avérer plus commun que ce qui est publiquement reconnu (Dörr, 2016).



Hanna Schevenels

Figure 2. Ligne du temps des événements marquants de l'histoire du journalisme automatisé.

1.4. Les États-Unis : berceau du robot-journalisme

Dans les sections suivantes, nous listerons les réalisations qui se sont distinguées dans l'histoire de la génération automatique de textes en journalisme. La *figure 2* (ci-dessus) constitue une ligne du temps résumant les événements clés abordés. Ainsi, l'automatisation des récits journalistiques a été lancée par deux compagnies américaines, Narrative Science et Automated Insights. Ces deux entreprises se partagent aujourd'hui le marché de la génération automatique de textes outre-Atlantique (Latar, 2018).

1.4.1. Narrative science

La compagnie commerciale Narrative Science est née d'un projet académique dans les laboratoires de la *Northwestern University's computer science and communication labs*, développé par les professeurs Larry Birnbaum et Kristian Hammond. Ce projet, intitulé *Stats Monkey*, avait pour objectif premier d'aider les journalistes au cours de leurs tâches quotidiennes. Il s'agissait de la première tentative commerciale visant à convertir automatiquement des faits en articles journalistiques (Latar, 2018 ; Dierickx, 2019). L'algorithme *Stats Monkey* a ainsi été programmé pour générer des résumés de matches de baseball grâce à des statistiques de jeu (Latar, 2018). Le baseball constituait un point de départ idéal au vu de la grande quantité de données et de statistiques accessibles (Graefe, 2016). À cette époque déjà, Larry Birnbaum et Kristian Hammond s'étaient montrés prudents en affirmant que leur objectif n'était pas de remplacer les journalistes humains, mais bien d'améliorer leur efficacité et leur productivité (Latar, 2018).

En 2010, le projet *Stats Monkey* a évolué en la start-up Narrative Science³⁰. Grâce à une équipe d'experts en *computer science* et en communication, l'entreprise a développé et breveté Quill, un nouvel algorithme basé sur l'IA. L'objectif majeur de l'équipe était de créer une approche révolutionnaire à la génération automatique de textes en langage naturel. L'algorithme Quill opère à trois niveaux : la réception de données, l'extraction d'informations pertinentes grâce à des algorithmes d'IA et la transformation de ces faits en articles sans intervention humaine. Grâce à ces technologies, en quelques secondes, Quill est à la fois capable de délivrer des faits, mais aussi des prédictions (Carter, 2013 ; Marr 2015). Kristian Hammond a notamment expliqué : « Quill's power lies in the fact that it is a synthesis

³⁰ Source : www.narrativescience.com

of data analytics, artificial intelligence and editorial expertise »³¹ (Carter, 2013). Par ailleurs, il a également affirmé que l’algorithme développé par son équipe est capable d’ajuster le ton des articles générés en fonction des profils digitaux de chaque client (Latar, 2018).

Quill est programmé pour assimiler les usages de langage du domaine qu’il couvre et ainsi écrire de manière appropriée en fonction du sujet d’un article. Par exemple, dans le cadre d’une expérience, lorsque Quill a été programmé pour rédiger des articles à propos de chaque restaurant d’une ville donnée, on lui a notamment appris à reconnaître les différents composants d’une critique dans le domaine de l’Horeca, tels que les notes, le niveau du service, la qualité de la nourriture et les appréciations des clients. Rapidement, Quill a développé les structures narratives adéquates à la situation et a su générer des articles critiques pertinents (Podolny, 2015). Par ailleurs, l’algorithme de Narrative Science inclut l’utilisation de métaphores, qui sont pourtant considérées comme des concepts très « humains » (Latar, 2018).

The image shows a screenshot of a Forbes article. At the top, the Forbes logo is on the left, and navigation links for 'Billionaires', 'Innovation', 'Leadership', 'Money', 'Business', 'Small Business', 'Lifestyle', 'Lists', and 'Advisor' are on the right. Below the navigation, the date 'Jul 22, 2015, 12:00pm EDT' is displayed. The main headline is 'Earnings for FLIR Systems Projected to Rise'. Below the headline is the Narrative Science logo and the text 'Narrative Science Partner Investing'. A warning icon and text state 'This article is more than 4 years old.' The article content is presented in a list format with social media icons (Facebook, Twitter, LinkedIn) on the left. The text discusses Wall Street's expectations for FLIR Systems' second quarter results, mentioning a consensus estimate of 38 cents per share, a rise from 33 cents per share a year ago. It also notes that the consensus estimate remains unchanged over the past month but has decreased from three months ago when it was 39 cents. For the fiscal year, analysts are expecting earnings of \$1.68 per share. Revenue is projected to be 2% above the year-earlier total of \$369.4 million at \$378.4 million for the quarter. For the year, revenue is projected to roll in at \$1.56 billion. A paragraph states that revenue was down in the first quarter after rising in the previous quarter. The 7% drop in the most recent quarter brought revenue down to \$344.5 million. The quarter before that, revenue rose 9%. A final paragraph notes that over the last four quarters, the company has seen an average of 41% year-over-year growth in income. The biggest rise came in the fourth quarter, with income growing more than twofold from the year-ago quarter.

Figure 3. Exemple d'article généré automatiquement par Narrative Science sur le site de Forbes le 22/07/2015³².

³¹ Le pouvoir de Quill réside dans le fait qu’il s’agit d’une synthèse d’analyse de données, d’IA et de compétences rédactionnelles (notre traduction).

³² Source : <https://www.forbes.com/sites/narrativescience/2015/07/22/earnings-for-flir-systems-projected-to-rise/>

En 2012, lorsque le journaliste Steven Levy demandait à Hammond quel serait, selon lui, le pourcentage d'articles rédigés par des robots dans quinze ans, ce dernier répondait : « More than 90 percent³³ » (Levy, 2012). Il affirmait également que dans vingt ans, il n'existerait pas un seul domaine que les algorithmes de Narrative Science ne seraient pas capables de couvrir (Levy, 2012). Plus impressionnant encore, dans une interview donnée au *New York Times*, il assurait qu'un ordinateur gagnerait un jour un prix Pulitzer. Il avait à l'époque estimé que cela se produirait avant 2016³⁴, ce qui avait notamment fait polémique (Levy, 2012). Selon lui, au fil du temps les logiciels produits par Narrative Science seront assez compétents pour générer du journalisme d'analyse et des articles de long format (Latar, 2018).

Parmi les clients de Narrative Science, on retrouve notamment l'éminent magazine économique *Forbes*, le réseau de télévision *Fox* ou encore des sites internet couvrant le sport local (Dierickx, 2019). À ce jour, Narrative Science travaille à l'amélioration de la qualité de ses articles en créant des analytiques plus profondes, en améliorant l'utilisation des métaphores et en cherchant à générer des analyses comparées entre divers sujets (Goldberg, 2013). Il est cependant pertinent de rappeler que les algorithmes de Quill reposent sur deux préalables. Ils dépendent de données existantes et ne pourront générer des articles que si, et seulement si, ces données sont disponibles et répondent à des questions formulées par les développeurs. Ainsi, Quill ne pourrait pas prendre l'initiative de générer un article seul sans avoir été programmé à cet effet (Latar, 2018).

Étant considéré comme l'un des pionniers du développement du journalisme automatisé, il est intéressant de s'intéresser aux prédictions publiées par l'entreprise Narrative Science sur son site internet en 2017. Elle y explique notamment que les ingénieurs des plateformes ajouteront bientôt aux articles générés le recours aux sentiments, l'utilisation d'un ton particulier ainsi que des indices visuels. L'un de leurs objectifs majeurs dans les années à venir sera donc de dépasser le cadre de l'IA faible et de développer des systèmes capables d'interagir avec d'autres systèmes. Les *data scientists* souhaitent ainsi se rapprocher le plus

³³ Plus de 90 pourcent (notre traduction).

³⁴ Il est pertinent de noter que les avis divergent quant à la prédiction de Hammond. Toutes les sociétés de génération automatisée de textes ne partagent pas son avis. Claude de Loupy, CEO de la société française Syllabs, affirmait notamment lors d'une interview donnée sur la revue des médias de l'INA : « Cette 'prophétie' disant qu'une machine gagnerait le Pulitzer dans 5 ans a été faite... il y a plus de 6 ans. C'était peut-être une profession de foi, peut-être un coup de pub. Quoiqu'il en soit, c'est ridicule et dommageable. C'est une méconnaissance totale de ce que récompense un prix Pulitzer » (Didier & Raynaud, 2018). Il soulignait cependant qu'il est « probable qu'un journaliste gagnera un jour le prix Pulitzer pour un travail impliquant une collaboration entre humain et machine » (Didier & Raynaud, 2018).

possible du fonctionnement du cerveau humain (IA forte). Enfin, l'entreprise estime que la pratique du journalisme humain sera menacée dès lors que les organisations médiatiques compareront leurs retours sur investissements lors de l'utilisation de robots-journalistes. Cela pourrait, d'après eux, conduire à la diminution du nombre de places disponibles sur le marché du journalisme, mais également comporter des risques concernant la liberté d'expression (Latar, 2018). Ces conséquences seront discutées dans le deuxième chapitre de ce travail.

1.4.2. Automated Insights

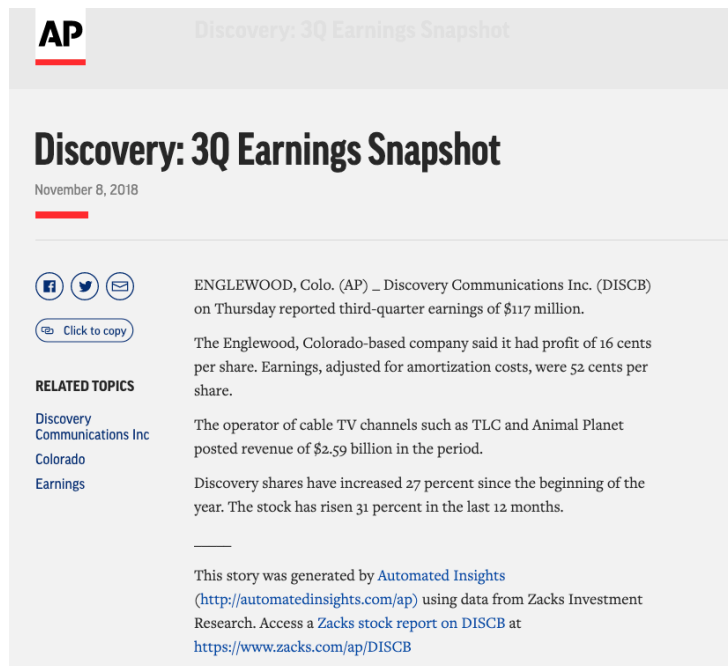
Après Narrative Science, Automated Insights se trouve être le second leader développeur de logiciels de génération automatique de textes aux États-Unis (Latar, 2018). La société est née sous le nom de StatSheet en juin 2007, fondée par Robbie Allen, diplômé du *Massachusetts Institute of Technology* (Dierickx, 2019). Tout comme Narrative Science, l'entreprise a commencé par se spécialiser dans la génération automatique d'articles dans le domaine du sport (Graefe, 2016). En septembre 2011, StatSheet a pris le nom de Narrative Science et a diversifié ses activités, annonçant qu'elle pouvait désormais générer automatiquement des comptes rendus financiers, des rapports marketing ou encore des annonces immobilières (Dierickx, 2019). Sur leur site internet, les développeurs comparent leurs algorithmes d'IA à un *data scientist* personnel qui analyserait d'énormes quantités de données et écrirait un article automatiquement, au détail près qu'il le ferait en quelques secondes et en temps réel, à l'échelle de plusieurs millions (Latar, 2018). James Kotecki, directeur des relations publiques d'Automated Insights a notamment affirmé que l'entreprise générerait des millions d'articles par semaine. Si besoin, le système pourrait créer jusqu'à 2000 articles par seconde (Miller, 2015).

L'algorithme d'Automated Insight peut être programmé pour créer des textes dans n'importe quel format journalistique : résumés, articles à points ou formats plus longs. Ces articles, rédigés en temps réel, peuvent ensuite être publiés à n'importe quelle échelle via n'importe quelle plateforme : e-mails, applications mobiles et tous les types de réseaux sociaux. Leurs algorithmes fonctionnent de façon similaire à ceux de Narrative Science (Latar, 2018). Le système de génération automatique rend des données structurées « plus humaines » et lisibles en identifiant des idées clés, en priorisant des informations basées sur leur contexte, notamment en fonction de leur degré de pertinence. Il construit ensuite un article

automatiquement dans une langue choisie et est capable de le publier en temps réel sur toutes les formes de sites médiatiques³⁵.

Automated Insights compte aujourd'hui parmi ses clients Microsoft, Bloomberg, MSN, Yahoo, *USA Today* ou encore l'agence de presse mondiale *Associated Press*³⁶. Depuis, 2014, *Associated Press* publie, en effet, tous les trimestres des centaines d'articles générés par les logiciels automatisés d'Automated Insight (Mc Farland, 2015). Lou Ferrara ancienne vice-présidente d'*Associated Press*, affirmait notamment à ce sujet :

We couldn't be happier. [...] When you look at the 170-year history of a company like the Associated Press, part of its success has been adapting and innovating in new media formats... We see this as part of that history³⁷ (McFarland, 2015, para.6).



The image shows a screenshot of a news article on the Associated Press (AP) website. The article is titled "Discovery: 3Q Earnings Snapshot" and is dated November 8, 2018. The article text reports that Discovery Communications Inc. (DISCB) reported third-quarter earnings of \$117 million. It also mentions that the company had a profit of 16 cents per share and that its revenue was \$2.59 billion. The article includes social media sharing icons (Facebook, Twitter, Email) and a "Click to copy" button. A "RELATED TOPICS" section lists "Discovery Communications Inc", "Colorado", and "Earnings". At the bottom, there is a disclosure: "This story was generated by Automated Insights (http://automatedinsights.com/ap) using data from Zacks Investment Research. Access a Zacks stock report on DISCB at https://www.zacks.com/ap/DISCB".

Figure 4. Exemple d'article généré automatiquement par Automated Insights sur le site d'Associated Press le 8 novembre 2018³⁸.

Grâce aux services d'Automated Insights, *Associated Press* publie désormais dix fois plus de comptes rendus concernant les revenus d'entreprises. Durant les derniers mois de 2014 par exemple, l'agence de presse a effectivement été capable de publier 3000 articles sur le sujet

³⁵ Source : www.automatedinsights.com

³⁶ Source : *ibidem*.

³⁷ Nous ne pourrions pas être plus heureux. [...] Lorsque l'on regarde les 170 ans d'histoire d'une entreprise comme *Associated Press*, une partie de son succès est liée à son adaptation et à ses innovations en termes de nouveaux formats... Nous considérons cela comme une partie de notre histoire (notre traduction).

³⁸ Source : <https://apnews.com/e4bbfaa752be29f67f3ef7649554df6c>

alors qu'elle ne pouvait auparavant en rédiger que 300 (McFarland, 2015). Par ailleurs, aujourd'hui, l'agence de presse mondiale a élargi le domaine de ses articles automatisés aux articles sportifs. La réaction des membres d'*Associated Press* s'est révélée très positive. En outre, le fait qu'une entreprise aussi importante et prestigieuse se lance dans le journalisme automatisé a été un élément déclencheur au sein de l'industrie journalistique, entraînant d'autres entreprises à envisager l'utilisation de ces nouvelles techniques (Graefe, 2016).

1.4.3. The Los Angeles Times

Au lieu de faire appel à une société extérieure, *The Los Angeles Time* a développé son propre algorithme générateur automatique de textes, Quakebot. Le journal californien est ainsi également considéré comme l'un des précurseurs en matière de robot-journalisme (Azar, 2014). Leur algorithme d'IA, composé par le journaliste et développeur Ken Schwencke, a été programmé pour générer un article lorsqu'un tremblement de terre se produit, grâce aux bases de données géologiques et aux alertes émises par le bureau géologique américain. Après relecture, cet article est automatiquement publié sur le site internet du journal (Latar, 2018). Quakebot est ainsi programmé pour générer un article comprenant toutes les informations basiques qu'un journaliste couvrirait, telle que l'heure du tremblement, sa localisation et son niveau de magnitude (Graefe, 2016). Le robot est utilisé au *LA Times* depuis 2011, mais ce n'est que deux ans plus tard qu'il s'est réellement fait connaître. Le 1^{er} février 2013, lorsqu'un tremblement de terre de magnitude 3.4 a fait trembler la Californie sur la côte de San Simeon, l'algorithme Quakebot a publié un article sur le tremblement, accompagné d'une carte indiquant son épicentre, en huit minutes seulement (Latar, 2018). Les dégâts de l'événement n'ont été que minimes, mais la technique de sa couverture médiatique a été retentissante puisque le premier article publié a été rédigé par le logiciel Quakebot. Lorsque les secousses de 2013 ont eu lieu, à 6h du matin, Ken Schwencke n'a eu qu'à relire et publier l'article, ce qui a permis au *LA Times* de relayer l'information en un temps-record par rapport aux autres médias (Azar, 2014).

Suite au succès de l'algorithme Quakebot, le *LA Times* a diversifié son offre d'articles rédigés par des générateurs automatiques de textes. Il utilise aujourd'hui plusieurs algorithmes différents, qui, à partir des bases de données du journal, rédigent de façon autonome des textes sur de multiples sujets. Les bases de données du journal sont nourries en permanence par les lecteurs, le gouvernement, les performances des marchés boursiers ou encore les

rapports de crime du service de police de Los Angeles. Un système d'alerte a également été mis en place concernant les arrestations de criminels hautement recherchés ou les homicides. Les algorithmes du *LA Times* sont donc à même de générer des articles couvrant ces différents domaines, et plus uniquement les catastrophes naturelles. Tous les logiciels sont programmés pour poser des questions qu'un journaliste d'expérience poserait habituellement dans une situation définie. Dans une affaire criminelle comme un homicide, par exemple, l'algorithme cherchera notamment des données concernant la personne qui a commis la plus grande offense en s'attardant sur les montants des cautions. Il s'agit là de se rapprocher le plus possible des informations qui intéressent les lecteurs (Marshall, 2013).

Los Angeles Times



Figure 5. Exemple d'article généré automatiquement sur le site du *LA Times* le 28/03/2013³⁹.

1.4.4. The Washington Post

Tout comme le *LA Times*, le journal américain *The Washington Post* a développé son propre algorithme générateur de textes, Heliograf. Selon Jeremy Gilbert, directeur de stratégies, Heliograf a été introduit dans la *newsroom* du *Washington Post* pour en faire un environnement productif pour les journalistes (Latar, 2018). Le software Heliograf identifie les données importantes parmi différentes bases de données, en crée des phrases correspondantes dans des *templates*, les assemble, puis publie différentes versions d'articles à travers différentes plateformes. Encore une fois, il est précisé que le but d'Heliograf n'est

³⁹ Source : <https://www.latimes.com/local/lanow/earthquake-38-quake-strikes-near-san-simeon-california-7w6zuc-story.html>

pas de remplacer les journalistes humains, mais bien de devenir un service complémentaire destiné à soulager les reporters en leur permettant de couvrir des événements moins répétitifs (Mullin, 2016). Heliograf a également été créé dans le but d'augmenter l'audience du *Washington Post*. Au lieu de viser une large audience avec un petit nombre d'articles qui demandent beaucoup de travail aux journalistes, Heliograph peut effectivement viser un grand nombre de petites audiences avec une multitude d'articles rédigés automatiquement sur des sujets divers et locaux (Keohane, 2017).

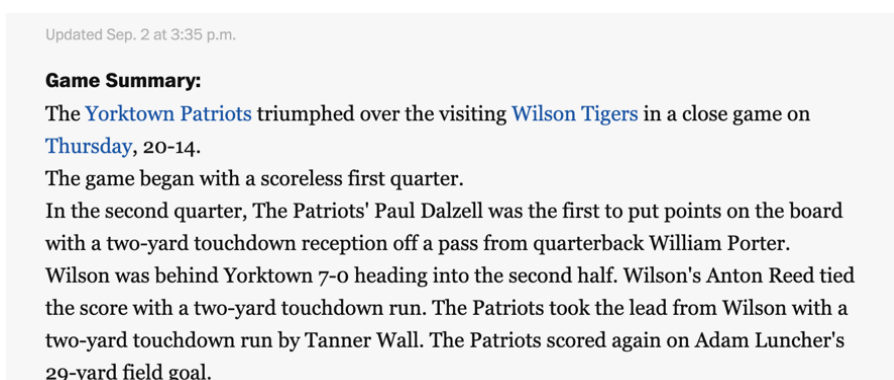


Figure 6. Exemple d'article généré automatiquement sur le site du *Washington Post* le 31/08/2017⁴⁰.

Heliograf a été utilisé pour la première fois au sein du *Washington Post* pour couvrir les Jeux olympiques de Rio, en 2016. Il a alors généré plus de 300 articles courts sur le déroulement des compétitions (Lambert, 2017). Il a également été utilisé pour rédiger plus de 600 tweets afin d'alimenter le compte twitter du journal (Dierickx, 2019). Au vu du succès de ce premier essai, le logiciel a été amélioré et ses compétences ont été étendues à d'autres domaines. Lors des élections américaines de 2016, Heliograf a été programmé pour couvrir l'événement : 500 articles ont été produits au total. Le journal a ainsi publié six fois plus d'articles que lors des élections de 2012 ; le *Washington Post* n'aurait jamais eu les capacités de faire rédiger un tel nombre d'articles par ses propres journalistes (Lambert, 2017). Aujourd'hui, le logiciel est également utilisé pour couvrir les matches de football des écoles secondaires (WashPostPR, 2017). Suite aux retours positifs, Jeremy Gilbert souhaite encore élargir les compétences d'Heliograf, notamment concernant la mise à jour automatique des articles ou la détection de tendances singulières au sein des prévisions météo, des résultats électoraux ou encore des rapports financiers (Lambert, 2017).

⁴⁰ Source : https://www.washingtonpost.com/allmetsports/2017-fall/games/football/87055/?itid=lk_inline_manual_3

1.4.5. Knowherenews

Le site internet américain knowherenews.com, créé en 2018, constitue un exemple particulier de l'utilisation des générateurs automatiques de textes en journalisme. Ses concepteurs ont développé la plateforme web *Knowherenews* notamment pour prouver que l'IA pouvait être utilisée de manière à produire des articles objectifs (DeGeurin, 2018). Les utilisateurs des réseaux sociaux se plaignent souvent que les plateformes ne leur montrent que du contenu déjà en accord avec leurs opinions. En journalisme, ce procédé est d'ailleurs largement utilisé aujourd'hui par les algorithmes, de manière à cibler le public et à maintenir l'attention des lecteurs (Targnion, 2019). L'équipe de *Knowherenews* a trouvé une solution pour contrer ce phénomène nommé la « bulle de filtres⁴¹ » : il s'agit de faire rédiger leurs articles par un logiciel d'IA spécialement programmé pour écrire objectivement (Ha, 2018). Sur la page d'accueil de leur site internet, ils indiquent « Join 250k readers and get straight facts – not opinions »⁴². Ils expliquent que pour protéger la société et la démocratie, chacun doit avoir la chance d'être informé correctement, peu importe ses opinions personnelles. Ils affirment également qu'à ce jour, les journalistes ont échoué dans cette tâche. Leur solution, pour remédier à cela, est de produire des articles neutres, exempts d'opinions humaines⁴³. Ils ont accompli cette ambition en créant un logiciel de génération automatique de textes conçu pour éliminer tout parti pris lors de la production d'articles (Ha, 2018). L'argument de l'objectivité des logiciels de génération automatique de textes a souvent été mis en avant dans le but de mettre en valeur le journalisme automatisé positivement. Nous aborderons ce point dans le deuxième chapitre de ce travail.

Pour générer un article, le logiciel d'IA de *Knowherenews* commence tout d'abord par rechercher quelles sont les informations les plus populaires du jour sur Internet. Il scanne à cette fin toute une série de sites médiatiques différents. Une fois une information importante choisie, le système se sert des sources les plus pertinentes pour générer un nouvel article,

⁴¹ La « bulle de filtres » est un concept médiatique développé par le militant américain Eli Pariser. Il s'agit d'un processus par lequel, sur internet, les algorithmes peuvent consigner une personne au cœur d'un univers informationnel clos. Le filtre se base notamment sur le profil de l'utilisateur, ses informations personnelles et ses habitudes de lecture. Il a souvent été affirmé que la « bulle de filtres » pouvait être responsable des polarisations idéologique qui ont lieu dans plusieurs pays (Bruns, 2019). *Knowherenews* vient à bout de cette « bulle de filtres » puisque chacun de leurs articles présente une version à priori neutre et objective d'une information.

⁴² Rejoignez notre communauté de 250 000 lecteurs et obtenez des faits – pas des opinions (notre traduction). Source : <https://knowherenews.com/>

⁴³ Source : <https://angel.co/company/knowwherehq/jobs>

exempt de subjectivité (DeGeurin, 2018). Le texte final est alors décliné en plusieurs versions : l'une neutre, l'une positive et l'une négative. Dans le cas d'articles politiques, le logiciel génère une version neutre, une version tendant vers la gauche politique et l'une présentant les faits en faveur de la droite politique (Ha, 2018). Ces articles sont parfois enrichis ou modifiés par les journalistes de la rédaction, permettant un travail de collaboration entre robot et humain. Selon les concepteurs, il s'agit de la seule manière de tendre vers la vérité, les lecteurs ayant ainsi accès à une information authentique et complète⁴⁴.

Apr 9

Senate fails to pass new COVID-19 relief bill

The U.S. Senate on Thursday failed to pass a bill providing an additional \$250 billion to help small businesses cope with the COVID-19 coronavirus and its economic fallout. After Democrats and Republicans were unable to agree on separate bills each party offered, Majority Leader Mitch McConnell (R-KY) told reporters that there will now be additional discussions between the leadership of both parties in the House and Senate...

- **Republicans** Democrat additions can come later
- **Democrats** Pass a \$500 billion bill now

MarketWatch, The Washington Times, Reuters, Vox, The Courier-Journal, Newsweek, CNBC, Reuters, International Business Time ...

Figure 7. Exemple de résumé d'article publié sur le site de Knowherenews.com le 9/04/2020⁴⁵.

Si le site internet permet bel et bien de donner au lecteur une vision relativement complète par rapport à une information, il est important de souligner que le logiciel créé par l'équipe de *Knowherenews* ne fonctionne que sur base d'autres articles ou productions d'entreprises médiatiques. Il nécessite dès lors impérativement l'intervention de journalistes humains, partiaux par définition, pour fonctionner. Par ailleurs, le logiciel ne peut jamais diffuser une information avant les autres journaux puisqu'il se contente de retravailler les sujets déjà relayés par d'autres médias. Il est néanmoins pertinent de noter que le logiciel de *Knowherenews* se distingue de ses prédécesseurs du fait qu'il ne génère pas uniquement des articles répétitifs issus de données chiffrées telles que les résultats sportifs. *Knowherenews* publie effectivement des articles sur les sujets médiatisés les plus populaires et houleux du jour. Ces informations concernent souvent des discordes politiques complexes. Grâce à l'utilisation de techniques de *Machine Learning*, le logiciel est capable de produire des contenus divers et parfois délicats, se détachant d'un modèle d'article très rigide et répétitif

⁴⁴ Source : ibidem.

⁴⁵ Source : <https://knowherenews.com/event/69dc0bdc-61c1-45d0-96b5-dbb12dfd1db4>

(DeGeurin, 2018). Cette façon particulière d'utiliser les techniques de génération automatique de textes n'est pas utilisée par les médias européens, qui, comme nous le verrons dans les pages qui suivent, ne produisent automatiquement que des articles routiniers à base de données simples et structurées.

1.5. Robot-journalisme en Asie

1.5.1. Robot-journaliste en trois dimensions

En 2007, une équipe d'ingénieurs japonais, menée par Rie Matsumoto à l'Université de Tokyo, a, pour ainsi dire, réalisé la métaphore du robot-journaliste. Cette dernière a effectivement créé un robot en trois dimensions, capable de se mêler à la foule, de mener des interviews, de prendre des photos et ensuite de composer un article automatiquement de façon autonome à partir des informations récoltées (Latar, 2015). Il s'agissait là de la première tentative de création d'un robot en trois dimensions pouvant réaliser toutes les étapes de la réalisation d'un récit journalistique de façon similaire à celle d'un humain. Le robot, construit sur la technologie Segway, était notamment capable de se déplacer dans les rues et donc d'assister à des événements. L'algorithme était conçu pour sélectionner une information en la classant en fonction de sa rareté et de sa pertinence. Une fois l'information choisie, il était programmé pour explorer le monde réel. Ensuite, les informations étaient transférées à un « classificateur d'informations » qui calculait leur « score informationnel ». Si le score dépassait un certain niveau, cela signifiait qu'elles pouvaient potentiellement intéresser les lecteurs et le logiciel générait alors un article automatiquement. En plus des logiciels de génération automatique de textes, le robot utilisait dès lors des formules mathématiques et statistiques tout au long du processus de récolte d'informations de façon à déterminer si les éléments découverts valaient la peine d'être publiés (Matsumoto et al., 2007).

1.5.2. Tencent

En 2015, le géant chinois de l'Internet Tencent a développé son premier algorithme de génération automatique de textes, Dreamwriter. Le 10 septembre de cette même année, le premier texte journalistique généré automatiquement était publié en Chine. L'article traitait de l'évolution de l'indice des prix à la consommation, comptait une centaine de mots et avait été rédigé en chinois en une minute (Dierickx, 2019). Dreamwriter est équipé de facultés de

calcul à très haute vitesse, d'une fonction de recherche d'informations et est capable de rédiger 1000 mots par minute (Latar, 2018). Encore une fois, il a été souligné que Dreamwriter n'a pas été développé dans le but de remplacer les journalistes humains, mais bien dans l'objectif de les libérer de tâches répétitives (Can, 2015). Aux Jeux olympiques de Rio, en 2016, durant les quinze jours d'événements sportifs, Dreamwriter a rédigé plus de 450 articles traitant principalement de la domination chinoise au cours des épreuves diverses. La publication a été rapide, apparaissant quelques minutes seulement après les événements. L'algorithme Dreamwriter est désormais également utilisé pour décrire des compétitions de jeux électroniques ; il produit des articles basés sur la façon dont les utilisateurs sont en train de jouer en temps réel (Latar, 2018).

1.6. La France : un acteur éminent

1.6.1. Deux sociétés de génération automatisée de textes

La France compte déjà deux sociétés de génération automatique de textes. Parmi elles, Syllabs, une start-up parisienne fondée en 2006 par Claude de Loupy et Helena Blancafort⁴⁶. La version initiale de leur moteur de rédaction a produit ses premiers textes en 2011, sur base de travaux réalisés durant cinq ans dans le domaine du traitement des langues. Depuis lors, leurs logiciels sont en constante évolution, de nombreuses fonctionnalités étant ajoutées au fil du temps. Syllabs se montre maintenant bien installée dans le paysage médiatique français et compte plus de quarante marques de médias en référence (entretien du 9/04/2020 avec Claude de Loupy). Parmi ses clients, on retrouve notamment France Télévision, *Les Échos*, *La Tribune*, *Slate*, *France Bleu*, *L'Express* et *Le Parisien* (Dierickx, 2019). La société vend également ses services au secteur de l'immobilier (Guy Hoquet), de la restauration (lafourchette), à la Bibliothèque nationale de France et à la SNCF, même si le journalisme reste leur domaine privilégié⁴⁷. Lors d'un entretien réalisé dans le cadre de ce travail, Claude de Loupy, CEO de Syllabs, expliquait : « Les médias sont notre principal secteur d'activité. Ils représentent plus de 50 % de notre chiffre d'affaires en 2019 » (entretien du 9/04/2020 avec Claude de Loupy).

⁴⁶ Source : <https://www.syllabs.com/fr/>

⁴⁷ Source : *ibidem*.

La deuxième société génératrice de textes française se nomme LabSense. Tout comme Syllabs, cette dernière propose l'automatisation de textes en plusieurs langues et dans de très nombreux domaines (immobilier, e-commerce, tourisme, etc.). Son portefeuille clients comprend notamment Leroy Merlin, Free ou encore Easy Voyage, TF1 et M6 Web⁴⁸. LabSense collabore également avec la société de presse écrite belge *SudPresse*, comme nous le détaillerons dans les pages qui suivent.

1.6.2. Le Monde : une première mondiale

Le journal français *Le Monde* travaille avec la société de génération automatique de textes Syllabs depuis quelques années déjà. Cette collaboration a fait énormément parler d'elle en mars 2015, lors des dernières élections départementales françaises. C'était alors Data2Content, un logiciel appartenant à Syllabs, qui avait été utilisé par *Le Monde* pour générer la grande majorité des articles publiés sur leur site internet (Lambert, 2017). Ainsi, plus de 30 000 articles ont été rédigés automatiquement, couvrant les résultats des élections de toutes les communes et cantons français, sans intervention humaine. Il s'agissait d'une première historique au sein du journalisme français, mais aussi plus généralement d'une première mondiale. Un logiciel de génération automatique de textes n'avait encore jamais été programmé pour réaliser un travail journalistique d'une telle ampleur (Glad, 2015).

RÉSULTATS RÉGIONALES 2015 - CHAMPS-SUR-MARNE

Résultats du second tour des élections régionales : commune de Champs-sur-Marne

Le second tour des élections régionales dans la région Île-de-France a donné lieu à une triangulaire entre M. Claude BARTOLONE (liste Union de la Gauche), Mme Valérie PECRESSE (liste Union de la Droite) et M. Wallerand DE SAINT JUST (liste Front National). Les scores réalisés par les candidats à Champs-sur-Marne sont les suivants : la liste Union de la Gauche de M. Claude BARTOLONE était en tête du scrutin, créditée de 49,33 % des suffrages. Les autres listes (liste Union de la Droite) et liste Front National) conduites par Mme Valérie PECRESSE et M. Wallerand DE SAINT JUST ont obtenu respectivement 34,35 % et 16,32 % des voix.

Le taux d'abstention pour ce second tour a été de 49,88 % dans cette commune.

Ces textes ont été écrits en collaboration avec Data2Content, une marque de la société Syllabs, à partir des données du ministère de l'intérieur et de l'Insee.

Figure 8. Exemple d'article généré automatiquement par Syllabs sur le site du monde.fr lors des élections départementales françaises de 2015⁴⁹.

Avant l'utilisation des algorithmes fournis par Syllabs, lors de chaque élection, le monde.fr présentait les résultats des votes de toutes les communes et cantons français sous forme de

⁴⁸ Source : <https://www.lab-sense.com/>

⁴⁹ Source : <https://abo.lemonde.fr/ile-de-france/seine-et-marne,77/champs-sur-marne,77083/elections/regionales-2015/>

données chiffrées brutes. Lors des élections de 2015, grâce au logiciel Data2Content, ces données ont systématiquement été accompagnées d'un texte rédigé automatiquement (Bronner, 2015). Ces articles, relativement courts, présentaient les pourcentages des voix obtenues par chaque candidat en lice ainsi que le taux d'abstention de la commune ou du canton concerné. Chacun des articles était suivi de la mention : « Ces textes ont été écrits en collaboration avec Data2Content, une marque de la société Syllabs » (Lambert, 2017). Lors d'une conférence au West Web Festival de 2018, Claude de Loupy soulignait notamment : « Le Monde nous publie sans nous relire » (de Loupy, 2018). Contrairement à l'exemple des articles rédigés par Quakebot au *Los Angeles Times*, qui étaient systématiquement relus par un journaliste avant publication (Glad, 2015), les articles du *Monde* générés par Syllabs ont effectivement tous été publiés automatiquement sans relecture humaine et sans aucune modification.

Pour Luc Bronner, directeur des rédactions du *Monde* : « Il nous paraissait intéressant d'expérimenter de nouveaux outils susceptibles d'apporter un nouveau service à nos lecteurs » (Bronner, 2015). L'objectif sous-jacent de cette expérience était d'innover et de proposer un nouveau mode de lecture aux utilisateurs. Par ailleurs, Luc Bronner rappelle que ces résultats d'élections n'auraient jamais pu être couverts de cette manière par des journalistes humains : « Les journalistes du *Monde* n'avaient de toute façon pas la capacité de produire 30 000 articles sur 30 000 communes en une nuit », explique-t-il. (Bronner, 2015). Il a également souligné que le but de cette opération ne visait aucunement à remplacer le travail des journalistes (Bronner, 2015). Lors d'une interview publiée sur la revue des médias de l'Institut National de l'Audiovisuel (INA), Claude de Loupy expliquait :

« Quel que soit le domaine, aucun système ne peut remplacer les capacités d'analyse, d'interview, de recoupement des journalistes. Nous pouvons retranscrire les résultats d'une élection dans une commune en le comparant à la moyenne nationale ou aux résultats passés de cette commune. [...] Mais nous ne pouvons interpréter les résultats au vu des déclarations des candidats de leur programme. Cela reste de l'apanage des humains » (Raynaud & Didier, 2018).

Les articles publiés sur *lemonde.fr* relèvent effectivement de la retranscription de faits et ne présentent aucune trace d'analyse. Ce travail spécifique a été pris en charge par les journalistes du *Monde*. En plus des articles générés par Syllabs, les élections ont effectivement mobilisé une septantaine de journalistes, dont une vingtaine d'envoyés spéciaux. Les articles rédigés par ces derniers ne relevaient pas des capacités du logiciel de

génération de textes et ont donc été publiés en parallèle aux articles rédigés par Data2Content (Bronner, 2015).

1.7. La Belgique : nouvel acteur

En Europe, sur la lancée de cette première mondiale du journal *Le Monde*, d'autres pays se sont investis plus activement dans la pratique du journalisme automatisé ou ont continué à développer leurs premiers essais. En Allemagne, depuis 2014, les procédés de génération automatique de textes étaient déjà bien installés, notamment par le journal *Berliner Morgenpost*, qui génère automatiquement des articles sur le niveau de particules fines à Berlin. Le *Weser Kurier* et le *Handelsblatt*, eux, utilisent la génération automatique de textes depuis 2015. Dernièrement, l'agence de presse norvégienne NTB, le média suédois *Östgöta Media* et les sites d'information du groupe suisse *Tamedia*, ont rejoint le mouvement (Dierickx, 2019). La Belgique, n'est maintenant plus en reste puisque plusieurs initiatives ont récemment été lancées dans le domaine du journalisme automatisé, notamment par *SudPresse* et *Mediafin*. On notera que ces deux sociétés de presse font toutes deux partie (entièrement pour *SudPresse* et à 50 % pour *Mediafin*) du groupe de presse Rossel, qui semble dès lors plus novateur en la matière que ses concurrents.

1.7.1. Première production automatisée d'informations en Belgique

En avril 2017 est née la première expérience de journalisme automatisé en Belgique. Dans le cadre de son doctorat en information et communication à l'Université Libre de Bruxelles, Laurence Dierickx a introduit un robot, nommé Bxl'air bot, au cœur de la rédaction du magazine d'actualités sociales *Alter Échos*. Bxl'air bot peut être défini comme une application qui collecte des données concernant la qualité de l'air en région bruxelloise en temps réel (Dierickx, 2019). Les données sont collectées par le robot via la plateforme CELINE, cellule interrégionale de l'environnement (Warsztacki, 2017). Une fois les données collectées, Bxl'air bot compare chaque résultat aux normes fixées par l'OMS et génère automatiquement des infographies ainsi qu'un bulletin quotidien sur le site de bxlairbot.be. Ces rapports, créés par le robot, étaient destinés à servir de brouillons d'articles pour les journalistes de la rédaction de l'*Alter Échos*. L'idée était que les journalistes puissent ensuite rédiger un dossier complet et pertinent concernant la pollution à Bruxelles (Dierickx, 2019). Ce projet, s'inscrivant au cœur d'une étude de long terme sur la qualité de l'air bruxellois, a

fait l'unanimité au cœur de la rédaction de *l'Alter Échos*. « Nous l'avons [Bxl'air bot] adopté sans l'ombre d'une hésitation », explique la journaliste Sandrine Warszacki (2017). Par ailleurs, en introduisant pour la première fois un robot dans une rédaction francophone belge, Laurence Dierickx souhaitait prouver son hypothèse que « l'on peut faire des machines des alliées » (Warszacki, 2017).



Figure 9. Exemple de bulletin généré par Bxl'air bot sur le site www.bxlairbot.be

Durant un an (du 1^{er} avril 2017 au 31 mars 2018), Bxl'air bot a donc enregistré et compilé des données et créé des ébauches de textes concernant la qualité de l'air à Bruxelles. L'intérêt du robot résidait dans le gain de temps d'une telle entreprise pour les journalistes. Ces derniers auraient été physiquement capables de réaliser les calculs que le logiciel exécutait, mais ne disposaient pas du temps nécessaire pour le faire. Le robot a donc permis l'automatisation d'une tâche lente et répétitive, mais aussi peu gratifiante (Gautier, 2018). L'expérience a été fructueuse, après un an, grâce aux infographies et aux textes générés par Bxl'air bot, les journalistes de *l'Alter Échos* ont pu réaliser un dossier d'analyse⁵⁰ documenté et réflexif sur le sujet de la pollution à Bruxelles (Gautier, 2018). Il est nécessaire de souligner que le projet s'inscrivait ici dans le cadre d'une collaboration entre robot et journalistes, les infographies et les textes générés par le logiciel ayant été enrichis de commentaires et

⁵⁰ Le dossier complet « Pollution l'air de rien » de *l'Alter Échos* est accessible à cette adresse : <https://www.alterechos.be/intro-dossier/>

d'analyses. Si le robot publiait automatiquement des rapports quotidiens sur son site internet indépendant, il a dès lors aussi permis aux journalistes de l'*Alter Échos* de nourrir une réflexion et une analyse de fond.

1.7.2. Sudpresse

SudPresse, leader de la presse quotidienne en Belgique francophone, travaille en collaboration avec la société de génération automatique de textes française LabSense. Depuis le mois de septembre 2019, ils publient sur leurs différents sites internet (*La Meuse*, *La Nouvelle Gazette*, *La Province*, *Nord Éclair* et *La Capitale*) des résumés des matches de football amateur rédigés par un logiciel sans aucune intervention humaine. Il s'agit d'une première au sein du journalisme belge ; *SudPresse* est aujourd'hui le seul média à publier des articles générés automatiquement sans les relire dans le domaine du sport en Belgique. Aucune information à propos de cette nouvelle collaboration n'a été trouvée dans la presse ou au cœur de lectures académiques. Par conséquent, les paragraphes suivants s'appuient sur un entretien réalisé dans le cadre de ce mémoire avec André Thiel, Directeur *Process & Development* chez *SudPresse*. Dans cette section, nous nous intéresserons principalement au développement des techniques d'IA au cœur de la société de presse écrite. Nous ne nous attarderons pas sur le contenu des articles, le chapitre 3 présentant une analyse détaillée de ceux-ci.

Sur un score de 2 à 1, le Royal Cornesse Football Club a gagné le match de la 26ème journée l'opposant au S.C. Saive à domicile...

Note importante : cet article a été rédigé sur base des données officielles de l'URBSFA

Le [Royal Cornesse Football Club](#) a accueilli le [S.C. Saive](#) pour un [match de Provinciale III série C \(Liège\)](#), au Complexe Bouhais, pour un match arbitré par Renaud Delbovier. Le Royal Cornesse Football Club a remporté la rencontre 2 à 1.

Le premier but de cette rencontre a été inscrit par Xavier Mostert à la 41ème. L'arbitre a sifflé la fin de la première mi-temps sur un score de 1-0. Le but d'Esteban Sorbian, inscrit à la 75ème, est venu remettre les deux équipes à égalité. Dorian Paulus a marqué sur penalty l'ultime but de la rencontre à la 88ème.

Figure 10. Exemple d'article généré par LabSense sur le site de Sudpresse le 8/03/2020⁵¹.

⁵¹ Source : <https://www.lameuse.be/531834/article/2020-03-08/provinciale-iii-serie-c-liege-succes-2-1-du-royal-cornesse-football-club-face-au>

SudPresse s'est lancé dans la génération automatique d'articles sous forme de collaboration, notamment avec la société parisienne LabSense, mais aussi avec la Fédération Royale belge de football (URBSFA) qui leur fournit une base de données fiable et détaillée concernant chaque détail des matches de football belge. André Thiel précise :

Il faut savoir qu'aujourd'hui, les feuilles de matches sont informatisées. C'est ça qui a permis, ces dernières années, d'obtenir assez vite, après la fin de la rencontre, les indicateurs principaux d'un match. À savoir : le résultat, qui a marqué à quelle minute, le nom de l'arbitre, les gens qui ont joué, les remplaçants, etc. (entretien du 30/03/2020 avec André Thiel).

Grâce à ces données structurées, le logiciel fourni par LabSense génère automatiquement des résumés simples des matches. Ces articles couvrent toutes les rencontres de football amateur dont les données sont répertoriées par l'Union belge de Football⁵². Cela correspond à la publication automatique de plus de 6 000 articles par week-end. André Thiel souligne la simplicité du processus d'IA utilisé :

Au fond, l'IA, dans ce cas-là, c'est essentiellement un tableau Excel qui est donné à manger à un convertisseur de textes. Et l'expertise des boîtes comme LabSense, elle est dans la sémantique et dans le fait de réussir à écrire des articles, de varier les tournures de phrases pour ne pas toujours écrire les mêmes articles sur tous les matches (entretien du 30/03/2020 avec André Thiel).

Ce processus permet au lecteur d'obtenir un texte cohérent et correct concernant le déroulement du jeu et les résultats. Par ailleurs, les articles sont enrichis par des infographies, également générées par des processus relevant de l'IA. Sous le chapeau de chaque article, il est systématiquement indiqué : « Note importante : cet article a été rédigé sur base des données officielles de l'URBSFA ». Cependant, il n'est pas mentionné clairement l'utilisation des logiciels de la société LabSense. Aucun indicateur ne permet effectivement au lecteur de prendre connaissance que le texte qu'il a sous les yeux n'a pas été rédigé par un journaliste humain. Le chapitre 2 de ce travail abordera notamment ce point éthique.

André Thiel souligne que les matches de football amateur n'étaient pas couverts par *SudPresse* avant la mise en place du partenariat avec LabSense : « C'est hors de portée, on n'arriverait clairement pas à couvrir plus de 6 000 articles par week-end » (entretien du 30/03/2020 avec André Thiel). Si ces articles ne pouvaient pas être rédigés par des

⁵² À noter que les résumés des matches de football professionnel ne sont pas rédigés par un logiciel d'IA. La génération automatique de textes n'est utilisée que dans le cadre des matches amateurs.

journalistes humains, l'automatisation a donc permis d'étendre le champ de couverture de *SudPresse*. Par ailleurs, André Thiel explique que les journalistes ne doivent pas se montrer inquiets par rapport à cette technologie, qui, selon lui, ne remplacera jamais le journalisme traditionnel. Le travail réalisé par le logiciel d'IA n'a pris la place de personne, explique-t-il : « On explore vraiment un nouvel espace, ça n'a pas changé du tout la manière dont les journées se passent pour nos journalistes. Ça a juste ajouté une couche » (entretien du 30/03/2020 avec André Thiel). Il explique aussi que le travail réalisé relève « de la pure data » et non d'un travail journalistique valorisant. Selon lui, les journalistes ne devraient pas avoir à réaliser ces tâches répétitives :

Aujourd'hui, il y a encore des journalistes qui passent leur temps avec leurs petites mains le dimanche soir à taper le nom de l'arbitre, qui a joué, dans des petites cases. Ça, c'est du travail inutile, un journaliste a bien mieux à faire que d'aller encoder de la data (entretien du 30/03/2020 avec André Thiel).

André Thiel explique notamment que les articles générés par LabSense peuvent être utilisés par les journalistes comme point de départ d'une analyse. Ces derniers peuvent modifier un article rédigé automatiquement et l'enrichir avec de la véritable valeur ajoutée journalistique, telle que des recherches d'informations ou des interviews. Selon André Thiel, l'IA peut ainsi aussi être valorisée comme une aide à la rédaction.

Il est intéressant de mentionner que les articles publiés grâce au logiciel d'IA ne sont pas relus avant publication sur le site internet de *SudPresse*. Le programme, maintenant validé par la rédaction, a été le sujet d'une très longue procédure de développement. On peut notamment ici, de nouveau, évoquer la théorie des champs de Bourdieu (1996), la société de génération de textes LabSense (acteur extérieur) ayant eu une influence sur le champ journalistique de *SudPresse*. Les deux champs ont été influencés l'un par l'autre et chacun a dû s'adapter aux règles de l'autre. André Thiel explique notamment que cette collaboration n'a pas impliqué que deux acteurs, mais bien quatre : SudPresse, LabSense, l'URBSFA et la société IDalgo (spécialisée dans la production d'infographies). La création des algorithmes a ainsi réuni quatre champs différents et a demandé un énorme travail d'adaptation, de validation des textes, des formulations et des infographies. Aujourd'hui, l'IA est réglée de façon optimale et produit des textes cohérents, dispensant les journalistes de les valider avant publication. André Thiel affirme d'ailleurs que le journal ne dispose pas des ressources suffisantes pour relire plusieurs milliers d'articles par semaine.

Le groupe *SudPresse* se montre très satisfait de cette première expérience de journalisme automatisé : « On est un groupe de presse régional donc on a vocation de faire vivre les communautés locales », explique André Thiel (entretien du 30/03/2020 avec André Thiel). Demetrio Scagliola, rédacteur en chef des éditions *SudPresse* souligne : « Mettre en lumière ces milliers de jeunes et de très jeunes sportifs et sportives, partout en Wallonie et à Bruxelles, c'est une mission prioritaire pour un média local comme le nôtre »⁵³. La couverture de matches de football amateur permet de consolider les liens étroits avec les différentes villes belges. Le Directeur *Process & Development* précise :

Ces articles-là, on ne peut pas dire qu'ils performent, qu'ils font des dizaines de milliers de vues. De toute façon, nous, des milliers de vues, on en a déjà largement assez. Sur notre site *SudInfo*, par exemple, on dépasse le million de visiteurs uniques par jour. Donc ce qu'on cherche plutôt à faire avec LabSense, c'est de commencer à faire vivre une communauté autour de ces articles-là, de s'adresser aux entraîneurs, aux patrons de clubs sportifs. On souhaite créer un microcosme autour de ces articles (entretien du 30/03/2020 avec André Thiel).

Le but premier de l'introduction des articles générés automatiquement n'était donc pas d'augmenter considérablement le trafic sur les différents sites internet du groupe, mais bien de fidéliser de nouveaux profils et de se rapprocher de la vie des régions. Aussi, *SudPresse* a reçu des retours très positifs de la part de l'URBSFA, qui est ravie qu'ils couvrent, même de façon automatisée, des compétitions qui n'avaient jusque-là jamais attiré l'attention médiatique.

À ce jour, l'automatisation n'est utilisée chez *SudPresse* que dans le cadre de la couverture des matches de football amateur. Si l'élargissement du système à d'autres sports serait assez facile à mettre en place, André Thiel explique que le groupe souhaite d'abord consolider le nouveau microcosme construit avant de commencer quoi que ce soit d'autre ; il s'agira de leur objectif majeur pour la saison prochaine. Néanmoins, parallèlement aux articles sportifs, ils ont également décidé de générer des textes concernant la météo locale. Un article par jour sera bientôt automatiquement rédigé pour chaque localité à partir de données structurées. Par ailleurs, dans le cadre des élections communales de 2018, le groupe générait déjà de courts textes concernant chaque commune à partir d'indicateurs tels que le prix des habitations, le niveau de vie moyen ou encore le nombre de chômeurs. Ils souhaitent désormais élargir ce concept de façon à publier un article actualisé par commune et par semaine en croisant les

⁵³ Source : <http://smonclubdefoot.be/>

différents indicateurs pour que les articles se montrent plus « journalistiques ». De manière générale, le groupe est également occupé à diversifier les formulations de phrases des textes générés automatiquement, de sorte à publier des articles qui se ressemblent moins. Cela permettra notamment d'éviter la mention « Duplicate Content »⁵⁴ du moteur de recherche Google et donc d'être mieux référencé lors de la recherche d'un utilisateur.

1.7.3. Mediafin

Le groupe de médias aux 26 700 abonnés, *Mediafin* (*De Tijd*, *l'Écho*), vient lui aussi, de se lancer dans un projet de journalisme automatisé. Ce dernier a déposé un dossier au Digital News Innovation Fund (DNI), soutenu par Google, et a ainsi obtenu un financement de 210 000 euros pour développer son projet⁵⁵. À nouveau, à défaut de sources académiques ou d'articles de presse sur le sujet, cette section est documentée sur base d'un entretien avec Nicola Becquet, chef du projet et responsable de la stratégie numérique de *L'Écho*.

En partenariat avec la société française Syllabs, *Mediafin* est en train de développer un logiciel, nommé Quotebot, programmé pour générer des articles sur l'actualité des marchés tout au long de la journée. Il est prévu que les articles soient déclinés en une version francophone pour *L'Écho* et une version néerlandophone pour *De Tijd*. L'actualité financière est un domaine où la génération automatique de textes est utilisée depuis longtemps déjà. La vitesse à laquelle l'information peut être publiée se trouve être un élément clé ; les robots prennent l'avantage par rapport aux humains quant à la vitesse d'écriture. L'agence de presse britannique *Reuters* est adepte des services d'automatisation depuis quelques années déjà (Graefe, 2016). Selon Reginald Chua, directeur exécutif des opérations éditoriales chez *Reuters*, dans le domaine de la finance, la génération automatique de textes se révèle une véritable nécessité. Il a notamment affirmé : « You can't compete if you don't automate ⁵⁶ » (Graefe, 2016, p.21). C'est dans ce cadre, et en suivant l'exemple d'autres entreprises de presse, que *Mediafin* a décidé de développer son projet d'automatisation. Le logiciel Quotebot, créé par Syllabs, aboutira en juillet 2020⁵⁷ et, selon Nicola Becquet, permettra à

⁵⁴ Le « Duplicate Content » est un concept qui s'applique aux pages de sites web relativement similaires. Google n'indexe alors qu'une seule de ces pages dans son moteur de recherche, pour éviter les doublons.

⁵⁵ Source : <https://newsinitiative.withgoogle.com/dnifund/insights/quotebot-editorial-led-robot-delivered-insight/>

⁵⁶ Vous ne pouvez pas rivaliser si vous n'utilisez pas la génération automatique de textes (notre traduction).

⁵⁷ Le projet a pris du retard suite à la crise sanitaire provoquée par l'épidémie de COVID-19 et sera certainement finalisé plus tard que prévu.

leurs journalistes de libérer du temps pour s'adonner à des activités demandant une véritable plus-value journalistique. Si l'automatisation dans le domaine de la finance est déjà utilisée par plusieurs entreprises, le projet lancé par *Mediafin* reste assez spécifique et n'a jamais été développé de cette manière en Europe.

Quotidiennement, les journalistes de *L'Écho* tiennent un *live blog* où sont publiés des articles suivis reflétant l'actualité des marchés tout au long de la journée. Ces articles font notamment l'état des bourses, de leur ouverture jusqu'à leur clôture. Deux journalistes se relaient chaque jour à cette tâche au sein de la rédaction. Selon Nicola Becquet, ce travail est fastidieux :

Il s'agit de réécrire chaque jour à la même heure un texte similaire, ou qui varie évidemment en fonction des chiffres, mais qui est assez répétitif et qui demande d'aller chercher des données sur plusieurs sites et plateformes différentes. Ça a une vraie plus-value pour les lecteurs, mais du point de vue du journaliste c'est fastidieux et répétitif (entretien du 2/04/2020 avec Nicola Becquet).

Le logiciel Quotebot a été développé de façon à remplacer les deux journalistes qui se relaient à cette tâche « ingrate ». Le but de l'opération est donc d'automatiser une tâche répétitive, de soulager les journalistes par rapport à ce travail et de leur permettre de se concentrer sur une activité plus stimulante et gratifiante, telle que l'analyse.

Le temps de publication des articles qui seront générés sur le *live blog* se révèle un autre point positif à prendre en compte, selon Nicola Becquet. Un journaliste qui doit collecter des données, rédiger un article, le publier et le relire prendra considérablement plus de temps qu'une IA, souligne-t-il. Le logiciel n'aura en effet besoin que d'une minute pour rédiger un article après avoir reçu les données. Quotebot pourrait, par ailleurs, être programmé pour publier un article encore plus rapidement, mais *Mediafin* souhaite dans un premier temps valider ce premier modèle.

Comme dans le cas de *SudPresse*, les articles générés par Quotebot seront publiés tels quels. Le groupe de médias compte cependant les relire avant publication lors des premiers mois de lancement du logiciel, de façon à s'assurer que tout fonctionne correctement. Nicola Becquet souligne que les textes ne seront pas enrichis par les journalistes de la rédaction, mais qu'ils pourront servir de base à une analyse ultérieure. À propos des articles générés automatiquement dans le domaine de la finance, Claude de Loupy, CEO de la société Syllabs, expliquait :

La machine ne pourra pas interpréter le fait que ces mouvements [du cours de la bourse] sont le fruit de telle ou telle déclaration du pdf de telle société cotée. C'est là où le journaliste peut intervenir en prenant la main sur certaines valeurs et en ajoutant son analyse d'expert à un texte qui ne comporte qu'une analyse chiffrée (Raynaud & Didier, 2018).

Nicola Becquet se montre bien conscient que les articles générés par Quotebot ne seront pas « de la grande littérature. L'objectif, c'est que ce ne soit pas juste des chiffres qui soient mis dans un tableau », mentionne-t-il. Ces textes générés automatiquement permettront aux journalistes de réaliser des analyses plus facilement, d'y consacrer plus de temps et d'obtenir un résultat plus abouti. Nicola Becquet souligne : « Ce qu'on souhaite, c'est mettre en valeur la plus-value journalistique [...] Tout ce qui est analyse, connaissance du contexte, style aussi ». Par ailleurs, *Mediafin* compte développer une transparence totale quant à l'utilisation du logiciel de génération automatique de textes. Chaque article rédigé par Quotebot sera suivi de la formule « Produit par Quotebot ». L'article renverra également vers un lien expliquant d'où viennent les données. « On assume totalement cette partie-là et c'est normal d'avoir une transparence par rapport aux lecteurs », explique Nicola Becquet.

Le projet a été bien accueilli au sein de la rédaction de *L'Écho*. Le responsable de la stratégie numérique ne l'a lancé qu'après en avoir discuté avec les journalistes, qui se sont montrés enthousiastes dès le départ. Nicola Becquet explique :

Au-delà des petites blagues du style « on va être remplacés », ils [les journalistes] voient bien la plus-value et la possibilité pour eux de se concentrer sur d'autres choses, qui sont plus intéressantes. C'est une expérimentation qui vient de la rédaction (entretien du 2/04/2020 avec Nicola Becquet).

Ainsi, ce sont les journalistes eux-mêmes qui ont aidé à créer les textes qui ont servi de modèles lors de l'élaboration du logiciel Quotebot. Outre quelques incompréhensions quant aux difficultés techniques liées au développement et à la longue durée du processus, l'accueil a été favorable.

En parallèle à ce projet d'articles financiers publiés automatiquement sur son *live blog*, *Mediafin* compte également utiliser la génération automatique de textes pour améliorer son système de « portefeuilles virtuels ». Sur le site internet de *L'Écho*, chaque internaute peut créer son propre portefeuille d'actions virtuelles, dont il choisit les propres valeurs comme s'il investissait réellement. Le souhait de *Mediafin* serait d'envoyer un bilan personnalisé rédigé automatiquement chaque matin à chaque utilisateur par e-mail. Si ces différents projets

se déroulent comme prévu, Nicola Becquet envisage notamment d'élargir les techniques de génération automatique de textes à d'autres domaines, tels que l'immobilier. Il souhaite néanmoins, avant tout, finaliser le projet en cours et s'assurer de son bon fonctionnement.

CHAPITRE 2

Questions clés et enjeux

2.1. Mise en contexte

L'évolution des techniques de génération automatique de textes a radicalement transformé les salles de presse de nombreuses rédactions (Ali & Hassoun, 2019). Par conséquent, elle a également redéfini le métier de journaliste et la place qu'il occupe au sein d'une entreprise médiatique (Kim, 2017). Wu, Tandoc, et Salmon (2019) affirment même que le journalisme automatisé a eu un impact considérable sur la société. Ce point de vue peut être considéré comme du déterminisme technologique⁵⁸, mais pas seulement, car il s'inscrit également dans le cadre d'explications plus nuancées et de schémas plus larges impliquant des contextes sociaux, politiques, culturels et économiques (van Dalen, 2012 ; Örnebring, 2010).

L'introduction des générateurs automatiques de textes en journalisme a provoqué différentes réactions, tant positives que négatives concernant l'évolution du métier. La vague d'avis optimistes, soulignant notamment le gain de temps que le journalisme automatisé ferait gagner aux journalistes, a vite été tempérée par des avis plus critiques, mais aussi par des évaluations plus réalistes de la pratique, de ses effets, son contenu et sa réception (Thurman, 2019). Parallèlement, de nombreuses réactions technophobes se sont fait entendre, désignant un profond rejet de la technologie, non seulement de ses usages, mais plus généralement de son essence même. Les personnes considérées comme « technophobes » associent les avancées technologiques à une aliénation et un démantèlement des liens sociaux les plus primaires (Andrieu, 2011). De manière générale, de nombreux experts ont émis des inquiétudes quant aux conséquences de l'évolution des IA sur l'essence de la vie humaine (Latar, 2018).

Claude de Loupy, CEO de la société de génération automatique de textes Syllabs, expliquait dans un entretien réalisé dans le cadre de ce mémoire : « La peur engendrée par ces

⁵⁸ Le déterminisme technologique est un courant social développé par le sociologue américain et théoricien de la communication Marshall McLuhan. Il suggère que les changements techniques ne sont pas influencés par la société et que l'évolution sociologique s'inspire des nouvelles techniques. En d'autres termes, les nouvelles technologies provoquent des changements sociaux et influencent la société, incarnant les rôles de sculpteurs de nos vies. De façon similaire, selon ce courant de pensée, les nouvelles techniques ont des conséquences importantes sur la société et sur les domaines dans lesquels ces dernières sont utilisées (Maigret, 2015).

technologies est logique : toute nouvelle technologie a généré des inquiétudes par rapport aux métiers qui ne l'utilisaient pas auparavant » (entretien du 9/04/2020 avec Claude de Loupy). Le CEO de Syllabs tente donc de relativiser le phénomène, les bouleversements technologiques ont effectivement souvent été accueillis avec méfiance à leurs débuts en journalisme (Latar, 2018). Il est également pertinent d'évoquer que, depuis plus d'un siècle, l'IA est utilisée par de nombreux scénaristes et romanciers pour créer des œuvres dystopiques où la machine cherche à prendre le dessus sur les humains. Les romans célèbres « 1984 », de George Orwell, et « Le dépeupleur », de Samuel Beckett, en sont des exemples. Dans ces fictions, les êtres artificiels attentent à la souveraineté des êtres humains qui finissent asservis par ces nouvelles technologies. Les humains perdent alors leurs spécificités et se font littéralement « gommer » par les êtres artificiels (Taillefer, 2009). En prenant en considération ces ouvrages, il n'est pas étonnant que l'IA ait soulevé des réactions parfois très réfractaires. Dans ce chapitre, et au vu des craintes développées suite à l'utilisation des générateurs automatiques de textes, nous aborderons les questions relatives à ses enjeux clés en presse écrite. Nous analyserons tout d'abord ses enjeux éthiques, déontologiques et légaux. Ensuite, nous poserons la question suivante : générateurs automatiques de textes en presse écrite, menace ou opportunité ?

2.2. Éthique et déontologie

2.2.1. Code de déontologie des IA ?

En 1993, Claude-Jean Bertrand, spécialiste français des questions d'éthique journalistique, établissait les M*A*R*S en Europe (Moyens d'Assurer la Responsabilité Sociale). Il y abordait notamment la responsabilité des médias à l'égard de la démocratie et soulignait l'importance primordiale de la déontologie journalistique :

Depuis une vingtaine d'années, j'étudie les questions de déontologie médiatique. Et il m'est devenu évident que d'elle dépendait la survie de l'humanité. Plus clairement : cette survie dépend de la généralisation de la démocratie ; il ne peut y avoir de démocratie sans liberté de presse ; cette liberté ne peut survivre sans déontologie (Bertrand, 1993, p.22).

Si à l'époque, Claude-Jean Bertrand n'abordait pas encore les développements technologiques dans le domaine du journalisme et leurs éventuels dangers pour la déontologie journalistique, il est évident que ces derniers ont bousculé la façon de travailler des journalistes et ont des implications déontologiques significatives. L'introduction des

algorithmes au cœur la production d'articles génère effectivement de nouvelles problématiques éthiques à plusieurs niveaux. Ces enjeux comprennent notamment l'origine des sources, l'objectivité des logiciels ou encore la transparence quant à l'utilisation des algorithmes (Dörr & Hollnbuchner, 2017).

La Déclaration des droits et des devoirs des journalistes, signée en 1971 à Munich (charte de Munich), adoptée par la Fédération Européenne des Journalistes (FEJ)⁵⁹, constitue aujourd'hui une référence européenne sur l'éthique journalistique (Glad, 2015). Cette dernière est composée de dix devoirs et de cinq droits⁶⁰. La question suivante se pose : les algorithmes d'IA respectent-ils cette charte ? Selon Vincent Glad (2015), journaliste chez *Libération*, le journalisme automatisé respecterait les huit premiers devoirs de la charte, mais serait plus tendancieux concernant les deux derniers :

- 9) Ne jamais confondre le métier de journaliste avec celui du publicitaire ou du propagandiste ; n'accepter aucune consigne, directe ou indirecte, des annonceurs ;
- 10) Refuser toute pression et n'accepter de directives rédactionnelles que des responsables de la rédaction⁶¹.

Les algorithmes d'IA étant sous l'influence directe des données qu'on leur soumet, il est effectivement prudent d'envisager que ces derniers génèrent des articles à partir de données ayant pu être fournies par des annonceurs ou des acteurs autres que les responsables de la rédaction. Le premier enjeu déontologique concerne donc l'exactitude et la provenance des données utilisées par les algorithmes. Des questions quant à l'origine des sources se posent également. Ces dernières doivent être pertinentes, correctes et objectives. Sur base de données erronées ou subjectives, un logiciel générera de toute évidence des articles incorrects ou subjectifs. Un autre enjeu concerne l'éventualité où des données seraient manquantes. Ce cas de figure pourrait également amener un logiciel à rédiger un article partial (Dörr & Hollnbuchner, 2017). Ces risques ne sont cependant pas abordés dans la charte de Munich, qui, signée dans les années septante, ne faisait pas encore mention de l'utilisation de logiciels de génération automatique de textes en journalisme.

⁵⁹ Créée en 1994, la FEJ constitue la plus grande organisation de journalistes en Europe. Elle représente les intérêts des syndicats et des associations de journalistes européennes. Elle se bat notamment pour les droits des journalistes en Europe, tels que la liberté d'expression et d'information. Son siège social se trouve à Bruxelles. Source : <https://europeanjournalists.org/fr/about-efj/>

⁶⁰ Source : <http://www.ajp.be/telechargements/droitsdevoirs.pdf>

⁶¹ Source : *ibidem*.

Si la charte de Munich reste un document européen officiel, la Belgique possède son propre code de déontologie journalistique. Ce dernier est composé de 28 articles, divisés en quatre parties, qui régissent les règles essentielles du journalisme⁶². Sur base de ce code, le Conseil de Déontologie Journalistique (CDJ) peut traiter des plaintes concernant les médias francophones et germanophones belges⁶³. Le code de déontologie aborde et traite avec attention les nouvelles technologies et plateformes utilisées par les différents médias aujourd'hui, mais, tout comme la charte de Munich, ne fait pas mention des générateurs automatiques de textes. De façon parallèle à la charte de Munich, certains articles du code de déontologie belge pourraient poser question, en particulier dans la quatrième partie du code, intitulée « respecter les droits des personnes ». L'article 25, par exemple, indique : « Les journalistes respectent la vie privée des personnes et ne révèlent aucune donnée personnelle qui ne soit pas pertinente au regard de l'intérêt général »⁶⁴. Comment un logiciel serait-il capable de déterminer ce qui ressort de la vie privée ? Est-il à même de définir ce qui ressort de l'intérêt général ? De la même façon, l'article 26 du code pose également question : « Les journalistes évitent l'intrusion dans la douleur des personnes et la diffusion d'informations et d'images attentatoires à la dignité humaine sauf ce qui est pertinent au regard de l'intérêt général »⁶⁵. Un logiciel peut-il discerner quelles informations sont jugées intrusives quant à la douleur d'une personne ? Nous pourrions aussi nous questionner sur la capacité des algorithmes à protéger leurs sources (Dörr & Hollnbuchner, 2017), ce qui constitue pourtant une règle fondamentale en journalisme. L'article 21 du code de déontologie journalistique pourrait ainsi aussi poser question.

Dörr et Hollnbuchner (2017) soulignent que l'on pourrait tirer la conclusion que la protection des sources ne s'applique pas réellement au journalisme automatisé, ce dernier étant généralement réservé aux sujets routiniers, qui ne demandent pas de protection des sources. Il est pertinent de relever que les domaines dans lesquels le journalisme automatisé est majoritairement utilisé sont effectivement moins souvent sujets à des problèmes éthiques ou déontologiques, car répétitifs (Latar, 2018). Selon Nicola Becquet, responsable de la stratégie numérique de *L'Écho*, on ne peut pas tout à fait parler d'éthique. Interrogé à ce sujet concernant le projet d'automatisation d'articles financiers chez *Mediafin*, il expliquait :

⁶² Source : <https://www.lecdj.be/wp-content/uploads/Code-deonto-MAJ-2017-avec-cover.pdf>

⁶³ Source : <https://www.lecdj.be/fr/>

⁶⁴ Source : <https://www.lecdj.be/wp-content/uploads/Code-deonto-MAJ-2017-avec-cover.pdf>, p.10

⁶⁵ Source : *ibidem*, p.11.

Est-ce qu'on peut parler d'éthique quand on produit des petits bulletins d'information automatisés à partir de chiffres ? Oui et non. [...] C'est une forme de mise en ligne et de rédaction spécifique pour mettre en forme des chiffres. Il n'y a pas de jugement, il y a des appréciations qui sont très techniques, voire scientifiques (entretien du 2/04/2020 avec Nicola Becquet).

Dans le cadre du projet de *Mediafin*, Nicola Becquet explique donc que rien ne touche concrètement à l'éthique. Il s'agit de traductions de chiffres en phrases, les algorithmes sont programmés pour générer des articles automatiquement en fonction de la façon dont les cours des bourses fluctuent. Par exemple, si le cours fluctue entre +0,1 et -0,1, l'algorithme est programmé pour affirmer que le cours stagne (entretien du 2/04/2020 avec Nicola Becquet). Le responsable de la stratégie numérique de L'Écho nuance cependant ses propos en expliquant que les modèles d'articles sur lesquels les algorithmes se basent ont été rédigés par des journalistes de la rédaction :

Ça a été fait par des journalistes derrière, donc la partie éthique et déontologique a été filtrée par des journalistes eux-mêmes, qui ont rédigé les modèles. On les entourait de précautions pour préciser d'où viennent les données, comment elles sont traitées, par qui, etc.

La FEJ adopte une position ouverte et positive quant à l'utilisation des générateurs automatiques de textes. Ricardo Gutiérrez, secrétaire général de la FEJ et membre du CDJ, souligne cependant que la déontologie journalistique doit rester d'application en toutes circonstances :

Les générateurs de textes, c'est très bien, tant [...] que les balises éthiques sont celles de la profession. On ne publie pas tout et n'importe quoi en tant que journaliste. Toutes les informations qu'on reçoit, on les trie, on les vérifie. Il y en a qu'on décide de ne pas publier, même si elles sont vraies, parce qu'elles pourraient mettre en danger la vie de quelqu'un par exemple ou qu'elles vont faire l'objet de représailles. Toutes ces normes, toutes ces balises éthiques, la machine ne les a pas forcément. Donc il faut vraiment qu'il y ait un contrôle humain sur la production (entretien du 30/03/2020 avec Ricardo Gutiérrez).

Ricardo Gutiérrez insiste sur le fait que la machine doit respecter toutes les règles déontologiques qu'un journaliste humain appliquerait, et ce, en toutes circonstances. Par ailleurs, il souligne que si les logiciels ne sont pas à même de vérifier le respect du code éthique, les articles générés automatiquement devraient être systématiquement soumis à une vérification humaine. Ombelet, Kuczerawy et Valcke (2015) vont dans le même sens. Ils affirment que le code de déontologie belge doit être respecté sans distinction entre journalistes et algorithmes. Les algorithmes ne doivent donc publier que des informations qui

sont de source sûre, par exemple. Ils doivent aussi pouvoir faire la distinction entre les faits, les analyses et les opinions. Ombet et al. (2015) soulignent également que les éditeurs doivent assigner une tâche de vérification à une intelligence humaine afin de contrôler l'exactitude des données et retirer toute information potentiellement incorrecte ou subjective. Cette vérification, cependant, déjà abordée dans le chapitre premier, s'avère difficile à mettre en place au vu du nombre important d'articles générés automatiquement. Dans le cadre des résumés de matches de football générés automatiquement sur le site internet de *SudPresse* par exemple, André Thiel nous expliquait qu'il serait inenvisageable de faire relire plus de 6 000 articles par semaine (entretien du 30/03/2020 avec André Thiel, Directeur *Process & Development* chez *SudPresse*). De la même façon, lors des élections départementales françaises de 2015, plus de 30 000 articles ont été générés automatiquement sur le site du *Monde* en une soirée (Bronner, 2015), ce qui rend la vérification humaine presque impossible.

Selon Ricardo Gutiérrez, il n'est néanmoins pas nécessaire de modifier le code de déontologie journalistique déjà existant :

Le code de déontologie, pour moi, s'applique déjà aux générateurs automatiques de textes. Le code de déontologie s'applique à toute production journalistique. On ne dit pas que ces productions doivent avoir été produites par un journaliste humain (entretien du 30/03/2020 avec Ricardo Gutiérrez).

Il affirme ainsi que la déontologie des IA est un « faux problème ». Un citoyen a effectivement le droit de déposer plainte au CDJ, peu importe qu'un article ait été produit par une machine. Selon lui, tout comme les humains, les logiciels doivent se plier aux règles déontologiques et prendre en compte la dimension éthique déjà existante. Paul Bradshaw (2016), journaliste anglais et auteur de *The Online Journalism Handbook*, est en accord avec les propos de Ricardo Gutiérrez. D'après lui, le journalisme automatisé suscite les mêmes questions déontologiques que les choix éditoriaux. Ce dernier affirme ainsi que les règles à respecter restent identiques, que les articles aient été écrits par un logiciel ou par un journaliste. Selon Gutiérrez et Bradshaw, les logiciels d'IA devraient donc être programmés de façon à se conformer aux règles de déontologie déjà établies. La réelle faisabilité de cette programmation n'a pas encore été analysée dans la littérature académique.

2.2.2. La transparence

Un autre enjeu éthique concerne la transparence quant à l'utilisation des logiciels de génération automatique de textes. La FEJ demande une transparence totale à ce sujet. Ricardo Gutiérrez indique :

Une autre balise que l'on demande, c'est que quand un texte est écrit par une machine, il faut que le lecteur le sache. [...] On ne doit pas utiliser ces machines à l'insu du lecteur (entretien du 30/03/2020 avec Ricardo Gutiérrez).

À ce jour, aucune loi n'a été établie concernant l'insert de cette mention. Comme expliqué dans le chapitre premier, des entreprises telles que *Le Monde* et *Associated Press* mentionnent clairement l'utilisation de l'IA sous chaque article généré automatiquement, alors que *SudPresse*, par exemple, ne l'indique pas. Graefe (2016) explique que la plupart des entreprises pourraient se montrer entièrement honnêtes quant à l'utilisation d'algorithmes. Cependant, beaucoup d'entre elles préfèrent ne pas l'être totalement, surtout si cela n'a pas été explicitement demandé (Graefe, 2016).

La transparence dans le domaine du journalisme est considérée comme un point central. L'idée a notamment été émise que le strict respect du principe de transparence serait la solution à la perte de crédibilité et de confiance dont les médias sont actuellement victimes (Hayes, Singer, & Ceppos, 2007 ; Diakopoulos & Koliska, 2017). Une honnêteté totale permettrait ainsi de créer des perceptions positives du journalisme (Plaisance, 2007) et rassurerait le public sur la justesse des informations reçues (Diakopoulos & Koliska, 2017). La transparence a même été définie par le technologue américain David Weinberger (2009) comme la « nouvelle objectivité ». Le concept d'objectivité a toujours été au centre des préoccupations journalistiques, permettant au lecteur d'avoir entière confiance en un article. Cependant, cet objectif s'est rapidement révélé inatteignable, un journaliste étant toujours influencé, consciemment ou non, par des partis pris. Selon Weinberger, le concept de transparence a remplacé l'objectif impossible de l'objectivité car il permet au lecteur de discerner la subjectivité d'un article (Weinberger, 2009). Si certains journalistes ont conservé le principe d'objectivité en tant que fil conducteur, les discussions à propos de la transparence n'ont fait qu'augmenter durant ces dernières années. Les recherches n'ont cependant pas encore prouvé que la pleine transparence au sein des entreprises médiatiques avait une influence notable sur la perception de crédibilité du public. Elle est néanmoins devenue une valeur éthique clé du domaine du journalisme (Diakopoulos & Koliska, 2017).

Dans une étude récente, Montal et Reich (2017) ont examiné la manière dont les articles générés automatiquement sont référencés sur les sites internet de douze entreprises médiatiques. Leur recherche montre que la plupart des médias font preuve d'un certain niveau de transparence par rapport à l'identité de l'auteur robotisé. Cependant, cette transparence se montre souvent partielle. L'auteur renvoie effectivement très fréquemment à la société de génération automatique de textes, l'entreprise médiatique elle-même ou le nom d'un algorithme, sans mentionner clairement que l'article a été rédigé par un robot. Selon Montal et Reich (2017), ces ambiguïtés suggèrent le besoin fondamental d'une politique de transparence totale quant à l'identité des auteurs.

Diakopoulos et Koliska (2017) ont mené une étude concernant la transparence dans le domaine du journalisme automatisé. Dans ce cadre, ils ont interrogé cinquante participants, comprenant à la fois des universitaires et des journalistes. Il est ressorti de cette expérience que la transparence se révélait essentielle à leurs yeux. Les participants ont effectivement affirmé vouloir qu'on leur indique l'identité de l'auteur d'un article, mais aussi l'identité des personnes qui ont créé et qui contrôlent les logiciels de génération automatique de textes. Selon eux, il devrait être signalé si les algorithmes ont été composés par l'entreprise médiatique elle-même ou par une société extérieure, cela ayant des conséquences quant à la subjectivité. Paradoxalement, des inquiétudes sont ressorties concernant l'utilité de ces indications, certains participants se sont demandé si elles ne surchargeraient pas les lecteurs d'informations superflues (Diakopoulos & Koliska, 2017). De façon similaire à Diakopoulos et Koliska, Thurman, Doerr et Kunert (2017) ont interrogé 22 journalistes et ont abordé la question de la transparence du journalisme automatisé. Les participants ont exprimé des réponses mitigées, mais la majorité d'entre eux s'est montrée fortement en faveur de la transparence. Des discussions ont eu lieu quant à la forme que cette transparence devait prendre. Une simple phrase indiquant qu'un article a été écrit par un logiciel est-elle suffisante ? Il n'y a pas eu de consensus par rapport à cette question. Certains participants ont émis des craintes quant à la surcharge d'informations pour les lecteurs, ce qui rejoint l'étude de Diakopoulos et Koliska (2017).

D'après Ombelet et al. (2015), la transparence se montre nécessaire, non seulement pour gagner la confiance du lecteur, mais aussi d'un point de vue légal, pour éviter toute action juridique dans le cas de contenu erroné. Selon eux, l'éditeur devrait se montrer complètement transparent et indiquer quels articles ont été rédigés par des algorithmes. Le besoin d'une

transparence quant à la signature est d'autant plus pertinent qu'il s'avère nécessaire (Azar, 2014). La plupart des études réalisées sur la réception des articles générés automatiquement démontrent effectivement que les lecteurs ne voient que peu de différences entre les textes générés automatiquement et ceux rédigés par des journalistes humains. Il ressort donc que la signature est essentielle si l'on souhaite que les lecteurs en soient conscients (Thurman, 2019 ; Clerwall, 2014).

Montal et Reich (2017) proposent un modèle à suivre lors de la publication d'articles rédigés par des logiciels d'IA. Ils suggèrent que chaque article mentionne clairement l'écriture automatisée et renvoie vers un lien de la société génératrice de textes. Ils préconisent également de préciser la source des données ainsi que la méthodologie utilisée par les algorithmes lors de la création de contenu. Dans le cas où un article a été révisé, corrigé ou enrichi par un journaliste, Montal et Reich suggèrent que l'article soit signé par le journaliste, mais précisent que les informations créées par l'algorithme devraient être identifiables. Ces suggestions, bien que réfléchies, ne sont actuellement pas des règles actées.

2.3. Devant la loi

En 1942, l'écrivain de science-fiction Isaac Asimov a conçu un ensemble de trois règles destinées à protéger les humains des éventuels dangers que les robots pourraient poser à la société (Latar, 2018). Les robots décrits dans ses romans devaient impérativement obéir à cet ensemble de lois (Asimov, 1983) :

- (1) Un robot ne peut porter préjudice à un être humain, ou, par inaction, laisser un être humain subir un préjudice.
- (2) Un robot doit obéir aux ordres qui lui sont donnés par des êtres humains, sauf dans le cas où de tels ordres entrent en conflit avec la Première Loi.
- (3) Un robot doit protéger sa propre existence aussi longtemps qu'une telle protection n'entre pas en conflit avec la Première ou la Seconde Loi (Goutefangea, 2016, p.2).

Plus tard, Asimov a ajouté une quatrième loi à celles créées précédemment, nommée la loi Zéro : « Un robot ne peut porter préjudice à l'humanité, ou, par inaction, laisser l'humanité subir un préjudice » (Goutefangea, 2016, p.10). Cette Loi Zéro précède toutes les autres en termes d'autorité (Latar, 2018). À l'origine, les robots évoqués par Asimov étaient des androïdes de science-fiction. Aujourd'hui, les développements de l'IA visent à imiter le fonctionnement du cerveau humain, ce qui donne une dimension très différente au concept

de robot. Cet objectif est évident en journalisme automatisé puisque les logiciels visent à remplacer les tâches habituellement réalisées par les journalistes humains. Ces algorithmes d'IA sont à l'extrême limite de la violation de la Loi Zéro d'Asimov. Ils ne peuvent effectivement pas être programmés de façon à garantir le respect sans condition des droits humains et la liberté de parole, cruciale à la survie sociale (Latar, 2018).

Il est important de rappeler que les lois d'Asimov sont issues du contexte de la fiction et qu'il s'agit davantage de principes éthiques que de véritables règles. Ces dernières constituent cependant un modèle de référence et sont reprises à de multiples reprises lorsqu'est évoquée une potentielle création de lois des robots. Si aucune loi concernant les utilisations de l'IA n'a encore été signée, le Conseil de l'Europe affirme qu'il existe un besoin essentiel de réglementation en la matière⁶⁶. Jan Kleijssen, Directeur de la Société de l'Information et de l'Action contre la criminalité du Conseil de l'Europe indique :

Dans le cadre du processus de réforme d'Elsinore du Conseil de l'Europe, le Secrétaire Général va proposer au Comité des ministres un agenda stratégique et à inscrire, à l'horizon 2028, la question de la régulation de l'IA comme l'un des défis majeurs afin de trouver un juste équilibre entre les bénéfices des progrès technologiques et la protection de nos valeurs fondamentales (Kleijssen, n.d.).

En septembre 2019, un comité ad hoc sur l'IA (CAHAI) a été créé au sein du Conseil de l'Europe. Ce dernier a été chargé par le Comité des Ministres d'examiner la faisabilité d'un cadre juridique pour le développement, la conception et l'utilisation de l'IA en suivant les normes du Conseil de l'Europe sur les droits de l'homme, la démocratie et l'État de droit (Burić, n.d.). Ce besoin de régulation est notamment demandé par la FEJ, indique Ricardo Gutiérrez :

On n'est pas contre cette évolution technologique-là [génération automatique de textes], pour autant qu'elle soit encadrée, c'est-à-dire qu'on ne doit pas laisser les robots prendre le pouvoir. Les générateurs de textes, c'est très bien tant que les logiciels sont sous contrôle humain (entretien du 30/03/2020 avec Ricardo Gutiérrez).

Le journalisme automatisé soulève des questions juridiques importantes et notamment des questions de responsabilités dans le cas où des articles générés automatiquement contiendraient des propos inappropriés, nuisibles ou illégaux. Lorsque des données sont

⁶⁶ Le Conseil de l'Europe a dédié un portail entier à l'IA sur son site internet, accessible à cette adresse : <https://www.coe.int/fr/web/artificial-intelligence/home>

incorrectes ou manquantes, il est aussi possible que les algorithmes génèrent des articles incorrects. De plus, les logiciels de génération automatique de textes ne saisissent pas toutes les nuances du langage humain, particulièrement les métaphores, l'humour ou la poésie, ce qui pourrait créer des incohérences de contenu (Ombelet et al., 2015). Puisque les algorithmes ne peuvent pas être tenus responsables de leurs erreurs, la responsabilité juridique retomberait alors logiquement sur une personne physique (Graefe, 2016).

Ombelet et al. (2015) ont analysé la question de la titularité du droit d'auteur dans le cas d'articles générés automatiquement. La Constitution belge déclare qu'un auteur est responsable de son article s'il est connu et domicilié en Belgique⁶⁷. Il sera alors pénalement et civilement responsable. Cependant, si l'auteur n'est pas connu, les autres acteurs de la chaîne de production seront tenus responsables. Jusqu'à présent, il n'existe aucune loi qui détermine qui est considéré comme auteur d'un article généré automatiquement. Ombelet et al. (2015) ont identifié quatre acteurs qui pourraient potentiellement être accusés en cas de contenu inapproprié : la compagnie qui a développé l'algorithme, la source qui fournit les données aux algorithmes, le rédacteur qui sélectionne les sources et supervise le travail des algorithmes et l'éditeur qui utilise les services de génération automatique de textes. Dans les circonstances où les données utilisées par les algorithmes sont fausses ou contiennent des informations confidentielles et délicates, la responsabilité pourrait en effet retomber sur l'éditeur, qui se doit de surveiller le contenu qui est publié sur ses canaux d'informations. Le journaliste qui supervise le travail et la source des données pourraient aussi être accusés de négligence. Dans ce cas, les éditeurs, qui ont le contrôle éditorial, seraient tenus responsables de juger de leur négligence. S'il s'avère que le problème vient du codage de l'algorithme lui-même, cependant, la responsabilité pourrait retomber sur les développeurs. Dans le cas particulier de propos diffamatoires, selon la loi belge, un plaignant ne peut déposer plainte que sur une personne physique. Il devrait alors prouver que l'article en question a été rédigé en collaboration avec un journaliste ou avec l'éditeur et que ce dernier avait l'intention malveillante de porter atteinte à quelqu'un ou à quelque chose. Ombelet et al. (2015)

⁶⁷ Source : https://www.senate.be/doc/const_fr.html

Article 25 de la Constitution belge :

« La presse est libre; la censure ne pourra jamais être établie; il ne peut être exigé de cautionnement des écrivains, éditeurs ou imprimeurs.

Lorsque l'auteur est connu et domicilié en Belgique, l'éditeur, l'imprimeur ou le distributeur ne peut être poursuivi. »

soulignent que, dans cette situation, le plaignant aura du mal à prouver l'évidence des intentions malveillantes.

Lewis, Sanders et Carmody (2019) relèvent que le scénario d'un article généré automatiquement qui serait diffamatoire semble relativement improbable au vu des sujets routiniers que la génération automatique de textes couvre actuellement. Cependant, ils affirment que les entreprises médiatiques doivent sérieusement considérer la responsabilité juridique à mesure qu'ils développent des robots-journalistes (Lewis et al., 2019). Ombelet et al. (2015) soulignent également le besoin essentiel de la création d'un cadre juridique. Jusqu'à ce qu'il soit clairement établi, ils expliquent que les entreprises médiatiques qui utilisent les techniques d'automatisation possèdent une responsabilité importante dans la protection de leurs lecteurs contre des informations fausses, nuisibles ou illégales. Dans le cas de la création des voitures sans conducteur, des compagnies telles que Volvo et Google ont déclaré entière responsabilité en cas d'accident (Gorzelany, 2015). Ombelet et al. (2015) affirment que les entreprises médiatiques devraient être conscientes qu'une responsabilité similaire pourrait leur être attribuée dans le cas d'un article qui poserait question au niveau juridique (Ombelet et al., 2015).

2.4. Menaces du journalisme automatisé

2.4.1. Disparition du journalisme humain

Les algorithmes de génération automatique de textes sont non seulement capables de créer des centaines d'articles, mais ils peuvent aussi le faire plus rapidement, de façon moins coûteuse et avec potentiellement moins d'erreurs qu'un journaliste humain. Ces logiciels sont en outre à même d'utiliser des données identiques pour créer des articles différents, dans plusieurs langues et selon de multiples angles, en les personnalisant selon les préférences individuelles des lecteurs. Enfin, les algorithmes ont le potentiel de générer des articles sur demande en réponse aux questions du public (Graefe 2016). Naturellement, ces développements ont engendré la crainte que les contenus automatisés n'éliminent un jour la profession des journalistes humains, ainsi que les salles de rédaction de manière plus générale. Alors que certains observateurs voient en l'automatisation de textes le potentiel d'augmenter la qualité du journal, les journalistes les plus pessimistes considèrent les robots-journalistes comme de véritables menaces à leur subsistance, à leur style de vie et à leur façon

de travailler (Graefe, 2016 ; Ali & Hassoun, 2019). La perte d'emploi est une crainte qui a été émise par beaucoup d'entre eux (Azar, 2014). La disparition des métiers humains se révèle effectivement un réel enjeu de l'utilisation du journalisme automatisé. Selon les futuristes les plus éminents, environ la moitié des métiers actuels risquent de disparaître suite à l'utilisation de l'IA (Rainie & Anderson, 2017).

La menace d'une disparition du journalisme humain a retenu l'attention ces dernières années et a souvent dominé le débat public (Graefe, 2016). Parmi la couverture médiatique, on notera le podcast *Planet Money*, diffusé par la radio américaine NPR, où un journaliste était en compétition avec un algorithme pour déterminer qui rédigerait le meilleur article (Smith, 2015). En 2015, *The New York Times* organisait un quiz qui permettait aux lecteurs de deviner si les textes affichés sur leurs écrans avaient été écrits par un journaliste humain ou par un algorithme (The New York Times, 2015). Si ces exemples relèvent du divertissement, ils montrent à quel point l'hypothèse du remplacement des journalistes par des robots est actuellement d'actualité dans les mentalités.

Selon Claude de Loupy, CEO de la société Syllabs, les algorithmes ne remplaceront pas les journalistes. Il s'est expliqué lors d'une conférence au West Web Festival de 2018 :

On nous dit toujours : « vous allez donc remplacer les journalistes ». [...] Ce que je dis aux journalistes si vous pensez ça, [c'est que] soit vous ne comprenez pas et vous n'avez pas connaissance de ce qu'on est capables de faire, et c'est évidemment ça. Soit vous avez une bien piètre opinion de votre métier (de Loupy, 2018).

Le CEO de Syllabs affirme que les algorithmes ne sont pas capables de « faire du journalisme », mais qu'ils sont uniquement limités à de la pure rédaction de textes (de Loupy, 2018). Il souligne également qu'au lieu de remplacer les journalistes, les logiciels sont destinés à collaborer avec eux, de façon à les compléter. Il estime que le travail des algorithmes ne rentre ni en concurrence avec le travail éditorial, ni avec le journalisme de terrain. Il les considère plutôt comme des « libérateurs de potentialité journalistique ». Ils permettraient ainsi aux journalistes de consacrer plus de temps à leur valeur ajoutée, aux interviews, à l'analyse, etc. (Raynaud & Didier, 2018). Selon ce point de vue, au lieu de se substituer au journalisme humain, l'introduction du journalisme automatisé permettrait au contraire de retrouver un journalisme de haute qualité. Dans la littérature académique, d'autres, comme Waleed Ali et Mohamed Hassoun (2019), confirment ce point de vue.

Au cours des entretiens menés dans le cadre de ce mémoire, les journalistes interrogés ont souvent abordé le concept de valeur ajoutée du journaliste, qui, selon eux, n'est pas remplaçable par une IA. Ils évoquaient notamment « l'intelligence humaine », « le sens de la réflexion », « l'émotion » et « la subjectivité ». De façon similaire, Claude de Loupy expliquait que les journalistes auraient très certainement toujours une valeur ajoutée par rapport aux logiciels, même dans le cas où l'on développerait un jour une IA forte :

Supposons qu'une IA ayant un niveau suffisant pour faire un travail de journaliste soit créée un jour. Je ne vois rien dans les laboratoires du monde actuellement qui me permette de penser qu'on va y arriver dans dix, vingt ou même 1000 ans, mais imaginons. Si une telle intelligence existe, elle ne fera que représenter un point de vue. Il est certainement impossible de pouvoir représenter toutes les individualités. Or, chaque individu apporte sa propre vision de monde. Je ne vois donc pas comment une machine pourrait supprimer la valeur ajoutée d'un individu de par le regard unique que celui-ci peut apporter (entretien du 9/04/2020 avec Claude de Loupy).

Selon Ken Schwencke, fondateur de l'algorithme du *LA Times*, les logiciels d'IA sont complémentaires aux journalistes. Ils leur permettent de ne pas perdre de temps sur des tâches répétitives et de se consacrer à l'analyse, à la création et aux interviews. Schwencke affirme donc également que le journalisme automatisé ne représente pas une menace. Au contraire, il participerait à rendre le journalisme plus intéressant (Azar, 2014). Plusieurs journalistes interrogés dans le cadre de ce travail partageaient cet avis. Marie Guérin, rédactrice en chef d'*ELLE Belgique*, se réjouissait notamment de pouvoir se consacrer à des tâches plus créatives :

Je sais que la profession considère ça comme une menace. À titre personnel, j'estime que si quelqu'un se sent menacé, c'est qu'il n'a pas foi en ses compétences. Moi, je crois vraiment en la technologie et j'estime que ce sera toujours une aide. Et si quelqu'un croit qu'il sera remplacé à 100 %, c'est que ce qu'il fait n'est pas suffisamment optimal. Moi, je suis persuadée qu'on aura toujours une plus-value et qu'on travaillera de concert. La technologie va juste nous apporter un confort. Par exemple, tous ces articles embêtants qu'on n'a pas envie de faire, ils les feront pour nous. Et on va pouvoir aller faire des portraits et des interviews créatives que la machine ne pourra pas faire. Donc moi, je vois plutôt ça comme un allié (entretien du 20/01/2020 avec Marie Guérin).

André Thiel, Directeur *Process & Development* chez *SudPresse*, tenait des propos similaires :

Il y a une phrase que j'aime bien, qui dit que si un journaliste pense qu'un robot peut lui prendre son travail, il a vraiment de quoi s'inquiéter et c'est peut-être qu'il a raison. Parce que ça veut dire que le journaliste pense lui-même qu'il n'a pas de valeur ajoutée. Moi, je

pars vraiment du principe que les générateurs de textes, c'est des tableaux Excel. Ce n'est pas plus séduisant qu'un tableau Excel (entretien du 30/03/2020 avec André Thiel).

Dans le cas où elle peut soulager les journalistes, la FEJ adopte, elle aussi, une position ouverte quant à l'utilisation des algorithmes de génération automatique de textes. Ricardo Gutiérrez expliquait :

Tant que ces outils sont des outils, c'est-à-dire qu'ils permettent aux journalistes de mieux faire leur travail ou qu'ils accomplissent des tâches ardues qui seraient parfois impossibles à effectuer pour un simple journaliste, notre position est assez ouverte. [...] Tant mieux si la machine prend en charge des tâches basiques, comme collecter les résultats de foot ou des résultats électoraux. [...] Est-ce qu'on a besoin d'un journaliste pour dire ça ? Non. Est-ce qu'une machine peut le faire beaucoup mieux et beaucoup plus rapidement qu'un journaliste ? Certainement, et qu'elle le fasse. Donc notre position est assez ouverte.

Ricardo Gutiérrez nuance tout de même ses propos en expliquant que les logiciels d'IA ne devraient pas remplacer les journalistes dans des registres délicats qui demandent un jugement humain ou qui engagent la ligne éditoriale d'un journal. Ils ne doivent, selon lui, être utilisés que pour accomplir des tâches ingrates et répétitives, ou impossibles à réaliser par un journaliste.

Même si les développeurs de logiciels de génération automatique de textes déclarent que leurs intentions ne sont pas de remplacer les journalistes, mais bien de les assister dans la tâche d'écriture, la crainte que ces algorithmes ne les remplacent un jour reste latente car les compagnies de médias sont toujours à la recherche de plus d'efficacité et de plus gros retours sur investissements (Latar, 2018). Ce malaise s'est ressenti auprès des journalistes interrogés dans le cadre de ce travail. Émilie Mascia, rédactrice en chef d'*ELLE à table Belgique*, évoquait notamment la peur qu'un robot soit rémunéré à sa place (entretien du 20/01/2020 avec Émilie Mascia). Julien Covolo, journaliste web à la RTBF, lui, soulignait que les logiciels d'IA peuvent être des outils très utiles pour les décharger de tâches répétitives et non gratifiantes et qu'ils peuvent notamment leur permettre de se consacrer à l'investigation. Cependant, la crainte d'être remplacé restait présente :

Ce qui me fait peur moi, c'est que ce soit utilisé pour remplacer un journaliste. Je crains, notamment, que dans les grands groupes de presse, on en arrive à cette situation. Oui, ça peut faire peur. Mais ce n'est pas la technologie en elle-même qui me fait peur, c'est ce que les gens pourraient en faire. Quelqu'un pourrait se dire qu'il n'a plus besoin de payer un journaliste pour écrire vu que c'est tout à fait factuel. [...] Ma crainte, ce serait qu'on dise

qu'à la place d'engager dix journalistes, on en engage huit et deux robots (entretien du 31/03/2020 avec Julien Covolo).

Ainsi, si plusieurs journalistes interrogés se sont montrés relativement confiants quant à la supériorité des capacités journalistiques humaines, certains redoutent néanmoins la manière dont les entreprises comptent utiliser les logiciels de génération automatique de textes et les conséquences que cela engendrera pour le marché de l'emploi.

Dans une étude publiée récemment, Andrey Miroshnichenko (2018) indique que les journalistes humains sont actuellement en compétition avec leurs collègues automatisés d'un point de vue à la fois quantitatif et qualitatif. D'après lui, les algorithmes ont déjà gagné la compétition d'un point de vue quantitatif puisqu'ils sont capables de publier un nombre beaucoup plus important d'articles en un minimum de temps. D'un point de vue qualitatif, en ce qui concerne le style d'écriture et les capacités journalistiques, il affirme que les journalistes humains perdront la « compétition » d'ici cinq à sept ans. Miroshnichenko (2018) explique néanmoins que si les journalistes humains finissent par conserver leurs emplois, les raisons n'en seront pas économiques, mais bien sociales. Dans de nombreuses entreprises, l'amélioration de l'efficacité de l'entreprise passe effectivement après le souci de protection des métiers humains, le socialisme prenant alors le dessus sur le capitalisme. Ce serait, selon Miroshnichenko, la seule raison pour laquelle le métier de journaliste serait toujours présent dans une dizaine d'années. Sur le sujet, cependant, comme les différents témoignages ci-dessus le montrent, les avis divergent, d'autres acteurs invoquant la plus-value journalistique que les algorithmes ne seront jamais capables d'atteindre. Le débat reste donc ouvert quant à la potentielle menace que l'IA représente pour les journalistes.

Quoi qu'il en soit, les algorithmes posent un challenge intéressant aux journalistes. Ils devront inévitablement rivaliser dans ce nouvel écosystème automatisé (Latar, 2018). D'après le chercheur allemand Andreas Graefe (2016), pour la couverture de sujets routiniers, les logiciels d'IA vont très certainement remplacer les journalistes. Pour maintenir la pratique du journalisme humain, il affirme qu'il serait souhaitable que ces derniers développent des qualités et des compétences que les robots-journalistes ne possèdent pas, telles que l'analyse approfondie ou les techniques d'investigation (Graefe, 2016). Afin d'éviter que les logiciels d'IA ne représentent une menace à leur subsistance, les journalistes devront effectivement apprendre à se dépasser et constamment chercher de nouvelles manières de se démarquer. En outre, ils devront se familiariser avec des méthodes

scientifiques d'analyse plus complexes (Latar, 2018). Latar (2018) préconise également qu'ils développent des connaissances multidisciplinaires larges dans les arts ou encore dans la philosophie, qui leur serviraient de base pour créer des contenus innovants et créatifs. Par ailleurs, Latar (2018) mentionne que les journalistes humains devront se familiariser avec tous les nouveaux outils digitaux qui leur sont accessibles et les utiliser pour générer des nouveaux formats, de sorte à unir leurs forces avec ces logiciels. Arjen van Dalen (2012) souligne que, plutôt que de rentrer en compétition avec les robots, pour survivre, les journalistes auront intérêt à se concentrer sur leurs atouts en évitant au plus possible les sujets routiniers, en développant une écriture plus créative et une couverture plus détaillée des événements. Le journalisme automatisé forcera ainsi les journalistes à exceller au sein des tâches où ils peuvent faire la différence (van Dalen, 2012). Dans un souci de redéfinition des rôles des journalistes humains à l'ère des algorithmes, plusieurs écoles de journalisme, dont la *Columbia School of Journalism*, organisent des ateliers sur les nouvelles pistes de travail à explorer (Latar, 2018).

Il est pertinent de mentionner que la potentielle disparition du journalisme a fait éclore des discussions quant aux implications qu'elle causerait au sein de nos sociétés démocratiques (Graefe, 2016), le journalisme étant identifié comme un rouage essentiel de la démocratie (Neveu, 2009). Selon le chercheur Noam Lemelshtrich Latar (2018), l'introduction des robots au cœur des fonctions traditionnelles des journalistes représente effectivement un réel danger pour les démocraties occidentales. Elle irait même jusqu'à mettre en danger la survie de ces dernières (Latar, 2018). La récente attaque à la liberté de la presse, couplée au manque de sensibilisation du public au rôle du journalisme libre laisse, selon lui, le terrain libre aux robots pour mettre en péril les fondements de la démocratie, ce qui constitue une violation claire de la loi Zéro d'Asimov (Latar, 2018). D'après Latar (2018), le seul moyen de protéger la démocratie des pouvoirs immenses de l'IA, serait d'appliquer une politique de transparence totale et de mettre en place une régulation de ses utilisations, comme évoquée dans les points précédents. Graefe (2016) souligne néanmoins que peu de recherches ont établi les éventuelles conséquences des algorithmes concernant la démocratie et il préconise dès lors que le sujet soit abordé plus en profondeur dans de futures études (Graefe, 2016).

2.4.2. Menaces pour la qualité du journalisme

Dans l'hypothèse où algorithmes et journalistes continuent à travailler de concert au sein des salles de rédaction, les journalistes émettent des craintes concernant l'évolution de la qualité des articles. Beaucoup d'entre eux considèrent le style d'écriture des logiciels d'IA comme une faiblesse (van Dalen, 2012) susceptible d'entraver l'image du journalisme. Au cours des entretiens menés dans le cadre de ce mémoire, plusieurs journalistes ont effectivement émis des doutes quant à la qualité du style d'écriture des logiciels d'IA. Émilie Mascia, rédactrice en chef d'*ELLE à table Belgique*, expliquait :

Les journalistes humains ont des qualités que les robots n'ont pas : le style d'écriture, l'aspect critique, la réflexion, la vérification des données. Dans le domaine *lifestyle*, ces éléments-là sont très importants. [...] On peut parler de tout ce qui est émotion. Qu'est-ce qui fait qu'un article marche mieux qu'un autre ? Ça va être en fonction du vécu d'une personne, de l'émotion. Je pense qu'un robot qui sortirait des articles sur des choses qui ne sont pas forcément en lien avec ce que les gens vivent, ça passerait peut-être très vite aux oubliettes (entretien du 20/01/2020 avec Émilie Mascia).

Selon Émilie Mascia, les algorithmes généreraient des articles que les lecteurs ne prendraient pas de plaisir à lire, et plus particulièrement encore dans le domaine du *lifestyle*. Nadia Chater, journaliste aux *Éditions L'Avenir* et coordinatrice de *Deuzio*, tenait des propos semblables :

Je ne suis pas sûre que les gens aient envie de lire de l'information générée automatiquement. Je crois que ce n'est même pas envisageable. [...] Franchement, je ne vois pas du tout ce que ça viendrait faire dans ce métier. Pour moi, ça ne s'adapte pas du tout au journalisme. Tout article doit être rédigé par un être humain. Sinon, ça n'a pas de sens. [...] Il faut une recherche, il faut un travail d'écriture (entretien du 30/03/2020 avec Nadia Chater).

Si ce commentaire se montre très réfractaire par rapport aux techniques de génération automatique de textes, il témoigne également d'un questionnement sur la qualité des articles générés automatiquement et leurs conséquences négatives sur le ressenti des lecteurs.

David Coppi, journaliste politique au *Soir*, émettait également des craintes concernant l'utilisation de la génération automatique de textes en journalisme. D'après lui, en supprimant la dimension subjective d'un journaliste, un article perd en qualité :

La différence c'est que le journaliste apporte sa dimension humaine et toute sa subjectivité : dans la relecture de l'information et dans le traitement, dans la restitution. [...] La subjectivité du journaliste peut être remplacée matériellement, mais en réalité, on perd le sens du journalisme, simplement. On doit garder une distance critique. La plus-value du journaliste,

c'est la subjectivité, ce qui est dans vos mains, dans vos yeux et dans votre façon d'écrire les choses, plus ou moins bien, de travers ou pas. Et normalement, ça, c'est intouchable, ou alors on sort du journalisme. On peut sortir du journalisme, c'est un choix (entretien du 31/03/2020 avec David Coppi).

En outre, même s'il affirme que les logiciels pourraient représenter une certaine aide, David Coppi émet de lourds doutes quant à la vérification des sources. Selon lui, sans une vérification des sources effectuée par un journaliste humain, un article perd en consistance, ce qui entrave la qualité du journalisme (entretien du 31/03/2020 avec David Coppi).

Nous pouvons ici évoquer l'existence d'un « contrat de lecture »⁶⁸, comme défini par Eliseo Verón (Gonzales, 1996), entre les lecteurs d'un journal et le journal lui-même. Les journalistes interrogés évoquent l'inquiétude que les articles générés automatiquement ne brisent ce contrat tacite, notamment en utilisant un style d'écriture différent et en étant exempts d'émotion et de subjectivité humaine. L'horizon d'attente, également défini par Verón, en serait alors modifié. David Coppi illustre cela en expliquant qu'en utilisant des algorithmes, « on sort du journalisme », et qu'il s'agit d'un « choix ». Selon Verón, cependant, le contrat de lecture permet de tisser un lien durable et privilégié avec le lecteur, ce qui est d'importance primordiale (Gonzales, 1996). Dans ce contexte, le journalisme automatisé représenterait une menace pour les entreprises médiatiques. Cet argument se révèle néanmoins discutable car les algorithmes peuvent être programmés de façon à s'adapter au style d'un journal et à générer des articles personnalisés en fonction des préférences d'un lecteur (Latar, 2018).

2.5. Opportunités du journalisme automatisé

2.5.1. Nouveaux horizons créés par les limites de l'IA

Une autre façon d'aborder la question des menaces de l'IA pour le journalisme humain serait de déterminer si la génération automatique de textes, de manière générale, a des limites et si

⁶⁸ Dans un entretien avec Pierre Gonzales, Eliseo Verón a situé le contrat de lecture « en tant que position de discours sous forme de stratégies énonciatives » (Gonzales, 1996, p.52). Il explique qu'il s'agit d'un dispositif d'énonciation, qui diffère selon les organes de presse, où est mise en forme la relation de lecture que propose le producteur du discours, en l'occurrence l'éditeur et l'équipe rédactionnelle, à ses lecteurs » (Gonzales, 1996, p.52). Il souligne que le contenu compte, mais que la façon de mettre en forme le contenu prime pour ce qui est de créer une relation particulière ou un lien privilégié, voire affectif, si possible durable, avec les lecteurs (Gonzales, 1996, p.52).

ces limites offrent de nouveaux horizons pour les journalistes. Selon Latar, (2018), les limitations actuelles de l'IA mettent effectivement en lumière les avantages des journalistes humains et créent des opportunités significatives pour eux. Plutôt que d'adopter un postulat technophobe, ce point de vue optimiste avance que la génération automatique de textes permet d'améliorer les techniques journalistiques et que les logiciels ne remplaceront pas les humains dans un futur proche (Ali & Hassoun, 2019). Pour développer cet argument, il est important d'établir plus clairement quelles sont les limites de ces algorithmes.

La première limite de la génération automatique de textes concerne les sources utilisées pour générer du contenu journalistique. Afin de construire une information quelconque, les logiciels demandent la mise à disposition de données de qualité structurées qui peuvent être lues par un logiciel informatique. En d'autres termes, il doit être possible de convertir ces données dans un tableur classique. En l'absence de ce type de données, aucun logiciel n'est, à ce jour, capable de rédiger un texte. C'est pour cette raison que le journalisme automatisé est massivement utilisé dans des domaines tels que la finance, le sport ou la météo, où les données sont abondantes et facilement accessibles (Graefe, 2016). Cette contrainte met en lumière la capacité des journalistes à couvrir tous les sujets pour lesquels des données ne sont pas disponibles et notamment tous les articles demandant des contacts interpersonnels, tels que l'investigation ou les interviews (Latar, 2018). La seconde limite des algorithmes d'IA concerne la fiabilité des bases de données sur lesquelles ils construisent leurs articles. Les logiciels sont effectivement dépendants de données qui peuvent comporter des erreurs ou des partis pris, ce qui pourrait générer la rédaction d'articles incorrects (Graefe, 2016). Cette limitation met cette fois en avant la faculté de vérification et de recoupement des sources des journalistes.

L'incapacité à expliquer des phénomènes nouveaux ou à produire une analyse représente une autre limitation importante des algorithmes. Ces derniers sont effectivement incapables d'établir des corrélations ou des chaînes de causalité entre les données qu'ils traitent (Graefe, 2016). En d'autres termes, ils ont la faculté d'expliquer comment un événement s'est déroulé, mais pas de développer pourquoi les choses se sont déroulées de la sorte. Les algorithmes sont ainsi limités à observer la société, sans l'analyser, contrairement aux journalistes (Lazer et al., 2014). Andreas Graefe (2016) souligne le fait que, détachées de leurs analyses, certaines données peuvent se montrer dénuées de sens, ce qui donne une nouvelle fois l'avantage aux journalistes. En outre, les algorithmes n'ont pas la capacité de rédiger des

opinions. Ils peuvent produire de nouvelles connaissances (après validation) mais ne sont pas capables de convertir des informations en propositions de loi ou en suggestions de changement, par exemple. Les algorithmes d'IA ne peuvent pas non plus « penser » en dehors d'un cadre conceptuel créé pour eux par des humains ou traverser différents cadres conceptuels (Latar, 2018). Les logiciels génèrent ainsi du contenu automatiquement selon un ensemble de règles établies et se montrent incapables d'innover. Leur utilisation est limitée à fournir des réponses à des problèmes clairement identifiés pour lesquels des données sont disponibles (Graefe, 2016).

Une des fonctions les plus importantes du journalisme consiste à alerter la société des risques potentiels associés aux changements environnementaux et sociaux qui constitueraient une menace à la survie du système social. Cette fonction ne peut pas être assurée par des algorithmes. Ces derniers peuvent aider à surveiller, à une échelle importante, des phénomènes qui pourraient menacer l'humanité, mais cette surveillance est elle-même accomplie suivant des indexes prédéfinis, créés selon des cadres conceptuels déjà existants. On ne peut donc pas espérer que les algorithmes comprennent et surveillent l'apparition de nouveaux développements alarmants. La capacité humaine à réaliser des liens entre des événements qui n'ont pas encore eu lieu auparavant leur est manquante. Il est également pertinent de mentionner que les algorithmes ne sont pas à même d'alerter la société à propos des dangers éventuels de l'IA elle-même, ou encore à propos de ses potentiels dangers physiologiques pour le cerveau humain. Il s'agit d'un problème essentiel que les futurs journalistes devront constamment aborder et qui ne pourra pas être laissé à la tâche des algorithmes (Latar, 2018).

La dernière limitation importante de la génération automatique de textes est liée à sa qualité d'écriture, inférieure à l'écriture humaine. À l'heure actuelle, les algorithmes sont restreints dans leur compréhension du langage naturel (NLP), spécifiquement concernant les informations contextuelles de l'ordre des idées, les métaphores, l'humour, le sarcasme et la poésie (van Dalen, 2012 ; Latar, 2018 ; Graefe, 2016). Les algorithmes d'IA ne peuvent pas appréhender ou exprimer des émotions humaines complexes ou des valeurs sociales, leur niveau actuel étant limité à l'IA faible. Leurs articles sont ainsi exempts de profondeur, d'empathie et ne montrent pas de compréhension des contextes culturels complexes (Latar, 2018). Terry Winograd (1991), éminent chercheur dans le domaine du NLP et de l'IA, a affirmé que la compréhension du langage naturel par les IA ne pouvait dépasser un niveau

bureaucratique, similaire à celui d'une personne qui n'aurait aucune empathie et qui agirait selon des règles mécaniques très strictes. Selon Winograd (1991), la génération automatique de texte actuelle se rapproche effectivement de la bureaucratie dans sa rigidité et son incapacité à s'adapter à des circonstances changeantes. Interprété dans le contexte du journalisme, ce point de vue sous-entend que les algorithmes sont incapables de rédiger des articles riches et profonds, contrairement aux journalistes.

Il a été établi que le journalisme automatisé pouvait paraître technique et ennuyeux aux yeux du public (Graefe, 2016). Des expériences ont en effet prouvé que les lecteurs préfèrent les articles rédigés par des journalistes humains plutôt que ceux générés automatiquement (Clerwall, 2014 ; Graefe et al., 2018), ce qui donne l'avantage aux humains sur ce point. Cependant, il est important de souligner que même si l'écriture automatisée paraît encore très technique aujourd'hui, dans les années à venir, il est plus que probable que les possibilités d'écriture s'améliorent et permettent aux algorithmes d'apporter plus de richesse et de profondeur à leurs productions (Graefe, 2016). Il existe un débat philosophique continu parmi les chercheurs concernant la question de l'évolution du NLP. Ils cherchent à déterminer si les machines seront, un jour, capables de comprendre sans restriction toute la richesse du langage naturel qui évolue constamment au sein de contextes culturels et sociaux complexes (Latar, 2018). Ce débat possède une résonance particulière dans le cadre du journalisme automatisé. Savoir si les algorithmes pourront un jour rédiger sans différenciation par rapport à un journaliste permettrait effectivement de déterminer s'ils représentent une menace pour le journalisme humain. À l'heure actuelle, les algorithmes n'en sont pas encore capables, ce qui met en lumière les qualités des journalistes.

Alors que certains chercheurs et journalistes les plus pessimistes considèrent les logiciels de génération automatique de textes comme de véritables menaces à leur subsistance, cette issue n'est, en définitive, pas inévitable. Les limitations inhérentes aux algorithmes d'IA offrent aujourd'hui aux journalistes humains des opportunités importantes. Pour en tirer parti, ils doivent cependant en prendre pleinement conscience et adapter leur mode opérationnel en fonction des capacités de ces nouveaux outils (Latar, 2018).

2.5.2. Créativité de l'IA

Les scientifiques et les philosophes débattent aujourd'hui à propos d'une question essentielle : l'IA sera-t-elle un jour capable d'égaliser ou de dépasser les capacités créatives du cerveau humain (Pereira, 2007) ? La réponse à cette question se montre d'importance primordiale pour le futur de l'ensemble des métiers, y compris celui de journaliste de presse écrite à l'époque de l'utilisation des générateurs automatiques de textes (Latar, 2018).

Le journalisme requiert une combinaison d'art et de travail scientifique. Le côté artistique s'exprime dans le souci permanent de trouver des idées innovantes, de nouveaux angles pour couvrir une information, mais aussi dans la recherche de nouvelles manières de susciter la curiosité du lecteur et d'augmenter les audiences. Le côté scientifique du travail journalistique, lui, implique l'utilisation d'outils d'analyse et de *data silos* où l'activité humaine est enregistrée et stockée. Aujourd'hui, l'IA fournit une alternative au travail scientifique en explorant et en identifiant des idées clés depuis d'énormes bases de données. L'IA fournit également aux journalistes des algorithmes capables de rédiger des textes narratifs basiques. Cependant, elle n'est pas capable de prendre en charge le côté créatif (Latar, 2018). Puisque les algorithmes d'IA ne peuvent pas « penser » en dehors des codes créés pour eux par leurs concepteurs, ils sont effectivement incapables d'atteindre le plus haut niveau de créativité. Le journalisme étant une activité artistique et créative, la réplique de l'ensemble des activités journalistiques par une IA s'avère donc, à ce jour, impossible (Miroshnichenko, 2018). Comme expliqué dans le point précédent, cette limitation inhérente à l'IA met en lumière des opportunités pour les journalistes humains. Dès lors, il est intéressant de déterminer si les IA seront, un jour, capables de développer cette créativité, en quoi les logiciels de génération automatique de textes représenteraient une menace future à la survie du journalisme humain. Dans le cas contraire, les journalistes maintiendraient leur supériorité par rapport à la machine.

Depuis les années 1950, les pères fondateurs de l'IA, John McCarthy, Herbert Simon, Marvin Minsky et leurs successeurs, ont prédit que tous les problèmes liés au développement de l'IA seraient résolus en une vingtaine d'années. Une septantaine d'années plus tard, malgré d'importants développements récents dans les domaines du *Deep Learning* et des réseaux neuronaux, les ingénieurs ont réalisé que le cerveau humain se révélait beaucoup plus complexe que ce qui n'avait été imaginé au départ (Knight, 2016). Selon les pères fondateurs,

cependant, il n'existe aucune limite à l'IA. Cette opinion est notamment soutenue aujourd'hui par Ray Kurzweil, futuriste éminent chez Google, et Daniel Kahneman, lauréat du prix Nobel d'économie en 2002 (Latar, 2018). Un courant de pensée rejette néanmoins les prédictions de Kurzweil et Kahneman selon lesquelles l'IA surpassera un jour le cerveau humain dans toutes les sphères d'activité. Cette école de pensée regroupe de remarquables chercheurs tels que John Searle, Margaret Boden et Nancy Andreasen, qui affirment que l'IA ne pourra jamais égaler ou surpasser la créativité du cerveau humain (Searle, 2013 ; Boden, 2004 ; Andreasen, 2014). Cette hypothèse est d'importance primordiale pour déterminer dans quelles mesures l'IA peut être créative et si elle pourra jamais répliquer la créativité des journalistes.

Le philosophe et linguiste américain John Searle se révèle l'une des figures emblématiques du courant rejetant les prédictions de Kurzweil et Kahneman (Latar, 2018). En 1980, il élaborait une expérience nommée la « chambre chinoise », destinée à réfuter le fait qu'une IA puisse être dotée d'un esprit et donc prouver l'inexistence d'une potentielle IA forte⁶⁹ (Damper, 2006). En 1990, suite à cette expérience, Searle publiait un article où il affirmait qu'une IA ne pourrait jamais entièrement imiter l'activité biologique du cerveau humain. Il y explique que les algorithmes d'IA ne peuvent pas « penser », ni développer d'émotion (Searle, 1990). Searle critique les philosophes d'IA qui avancent que l'esprit est une entité séparée du cerveau biologique, qui pourrait donc être entièrement programmée. Selon lui, la conscience et les autres phénomènes mentaux tels que les sentiments, la douleur, la soif, l'odeur et l'amour sont issus de processus neurobiologiques complexes du cerveau humain. Il explique que l'IA ne fait que manipuler des symboles et qu'elle se montre donc inapte à reproduire les processus neurophysiologiques du cerveau tels que la conscience. Au cours d'un *TED talk*, Searle a conclu que l'on ne pourrait jamais construire de machine consciente, la conscience ne faisant pas partie du monde physique, mais bien du monde spirituel, qui n'est pas du domaine de l'IA (Searle, 2013).

De façon similaire à Searle, Margaret Boden (2004), chercheuse en sciences cognitives, souligne l'inexistence d'une IA forte. Elle précise néanmoins l'argument de Searle dans le cadre plus précis du potentiel créatif de l'IA. Cette dernière certifie que le cerveau artificiel ne sera jamais capable d'atteindre le plus haut niveau de créativité du cerveau humain.

⁶⁹ L'expérience de la chambre chinoise visait également à démontrer l'incapacité du célèbre test de Turing à déterminer si une IA est capable de simuler les processus de pensée du cerveau humain (Damper, 2006).

Aujourd'hui, les algorithmes d'IA sont utilisés dans différents domaines créatifs, comprenant les arts ou encore la composition de musique (Latar, 2018), mais Boden ne considère pas ces algorithmes comme créatifs. Nous pouvons notamment citer l'exemple de l'algorithme AARON, développé par Harold Cohen en 1973, programmé pour peindre. Ce dernier est capable de créer des tableaux dans un style qui lui est propre ou dans le style de grands peintres connus (Cohen, 2016). Dans le même esprit, David Cope a développé un algorithme capable de composer des chansons sur base de styles déjà établis (Gioti, 2020). En 2016, Google a annoncé le lancement du projet Magenta dans le but de générer de la musique, des histoires et de l'art de façon automatique (Metz, 2017). Selon Boden (1998), ces exemples d'algorithmes ne peuvent pas réellement être qualifiés de créatifs car ils ne font que générer de nouvelles combinaisons d'idées pour lesquelles ils ont été programmés.

D'après Boden (1998), la créativité ne se résume pas uniquement à la combinaison de données, mais comporte deux autres éléments importants : l'aspect exploratoire et l'aspect transformationnel. Boden reconnaît que les algorithmes tels qu'AARON impliquent une dimension cognitive de la créativité, mais elle explique qu'elle se trouve limitée à reproduire la créativité humaine, et non pas à innover au sens premier du terme. Cette conséquence est liée à la difficulté des IA à simuler la richesse de la mémoire associative humaine, à identifier nos valeurs et à les exprimer dans une forme informatique. L'innovation viendrait des humains qui ont codé les algorithmes, et non pas des algorithmes eux-mêmes. Selon Boden, l'IA ne peut dès lors pas être considérée comme créative, il lui manque les aspects exploratoires et transformationnels essentiels à la créativité. En guise d'exemple, l'algorithme AARON ne peut pas réfléchir à ses propres productions ni les ajuster pour les rendre meilleures. En outre, le fait qu'AARON ne puisse pas être créatif en dehors des frontières conceptuelles définies par Cohen amène Boden à conclure qu'on ne peut pas espérer que de réelles ruptures ou innovations émanent de lui. Ses tableaux ne seront jamais que des mélanges entre des styles déjà établis (Boden, 1998). Cette conclusion a des implications significatives pour le journalisme puisqu'elle implique que les algorithmes ne seront jamais capables d'innover de façon similaire aux journalistes, qui, eux, pourront créer de nouvelles plateformes et de nouveaux modèles d'écriture pour concurrencer l'IA. Douglas Eck, du Google Magenta Project tient des propos similaires à Boden. Il reconnaît que la créativité des humains est supérieure à celle de l'IA, mais il y ajoute une nuance. Il souligne qu'une personne qui souhaiterait être hautement créative à un niveau transformationnel aujourd'hui devrait coupler sa créativité humaine à des techniques d'IA (Metz, 2017), ce qui

impliquerait une entraide entre robot et humain. Cette collaboration peut également être envisagée dans le domaine du journalisme.

Nancy Andreasen (2014), éminente neuroscientifique américaine, développe également l'idée que les algorithmes d'IA sont limités dans les tâches qui impliquent de la créativité. Son approche, faisant référence au QI, est cependant différente de celle de Boden. Andreasen a mené des expériences, réalisé des interviews et des scanners cérébraux pour découvrir quelle était la véritable essence de la créativité. Ces expériences ont montré que la créativité n'est pas directement liée au niveau de QI. Un QI élevé n'implique effectivement pas de hauts niveaux de réalisations créatives. La créativité est directement liée au subconscient, à l'intuition, au rêve, à la spontanéité, ou encore à l'intelligence émotionnelle. Une créativité élevée se produit lorsque le cerveau réalise des associations libres et pense de façon irrationnelle, ce qui n'est pas lié au QI et n'est pas non plus une faculté de l'IA. De façon opposée au cerveau humain, les algorithmes d'IA sont basés sur des processus de prise de décision rationnelle : les neurones artificiels font circuler une information aux autres neurones uniquement si cette information réduit le niveau d'incertitude. Selon Andreasen, ce mode de pensée rationnel empêcherait le développement d'une réelle créativité des algorithmes (Andreasen, 2014). Elle en conclut que, puisque les processus inconscients forment une partie très importante de la créativité, l'IA est incapable de développer un niveau de créativité similaire à celui des humains (Latar, 2018).

Si l'on rejette les convictions du groupe mené par Kurzweil et Kahneman et que l'on accepte la façon de penser de Searle, Boden et Andreasen, le cerveau humain aurait ainsi un avantage majeur pour ce qui est de la haute créativité. Il s'agirait d'une bonne nouvelle pour les journalistes, qui conserveraient une supériorité importante par rapport aux logiciels d'IA. L'introduction du journalisme automatisé pourrait, dans ce cas, être considérée comme une opportunité pour les journalistes de développer leurs aptitudes créatives.

2.5.3. Chiens de garde de la démocratie

Plusieurs chercheurs considèrent les générateurs automatiques de textes en presse écrite comme une opportunité pour les journalistes de retrouver leur rôle de chien de garde de la

démocratie⁷⁰ (Thurman, 2019). La fonction de chien de garde des journalistes, ou, comme l'a défini Edmund Burke, le rôle de quatrième pouvoir (Mc Nair, 2005), est identifié comme un moyen d'alerter les citoyens des potentiels abus de pouvoir d'un roi, d'un président ou d'un Premier ministre. C'est également une manière de faire rendre des comptes aux dirigeants, de dévoiler des méfaits et de rendre visibles des tendances sociales déviantes. Il s'agit de donner aux citoyens les informations dont ils ont besoin pour réaliser des choix importants (Thurman, 2019). Aujourd'hui, ce rôle est cependant soumis à de nombreuses épreuves et les journalistes expriment le regret de ne plus pouvoir l'exercer correctement. Le journalisme automatisé est ainsi considéré par certains comme un outil permettant de libérer les journalistes d'un travail peu gratifiant de découverte et d'obtention de faits, leur laissant alors plus de temps pour la vérification des sources ou l'analyse de temps long et leur permettant de se concentrer sur leur rôle premier (Flew et al., 2011).

Julien Covolo, journaliste web à la RTBF, expliquait ne pas toujours avoir l'impression de faire du « vrai journalisme ». D'après lui, le journalisme automatisé pourrait représenter une solution à ce problème :

Au niveau des générateurs de textes automatiques, je trouve que c'est une bonne chose si ça permet de décharger le journaliste d'une partie du travail qui pourrait être jugé comme ingrat. Je pense notamment à ce qu'on appelle le « bâtonnage » de dépêches, donc le fait de copier-coller des dépêches, de mettre des images. J'ai fait ça pendant un an, je n'ai pas l'impression de faire du journalisme quand je fais ça (entretien du 31/03/2020 avec Julien Covolo).

Les journalistes optimistes considèrent donc les générateurs automatiques de textes comme des outils qui les libéreront de tâches rébarbatives. Cela leur permettrait notamment de se concentrer sur l'investigation, leur talent narratif, ainsi que leur rôle de quatrième pouvoir (Latar, 2018).

Au nom de la FEJ, Ricardo Gutiérrez soutient ce point de vue. Interrogé sur les potentielles menaces que représente le journaliste automatisé, il répondait :

Non, au contraire, c'est une opportunité. Faire faire à des machines des tâches ingrates, c'est justement revaloriser le journaliste dans sa mission essentielle, qui est d'être au service du

⁷⁰ Dans la réalité des faits, cependant, les journalistes sont souvent critiqués pour être très proches des personnalités politiques et des groupes privilégiés (Milosavljević & Vobič, 2019). Cet argument est notamment développé dans le film *Les nouveaux chiens de garde*, réalisé par Gilles Balbastre et Yannick Kergoat. Le film pointe du doigt les relations qu'entretiennent de nombreux journalistes français connus avec des personnalités politiques et financières influentes. Il abat le mythe d'une presse qui se revendique objective, indépendante et pluraliste (Balbastre & Kergoat, 2011).

citoyen, d'être un contre-pouvoir. Défendre les positions et les intérêts des citoyens face à tout qui détient pouvoir dans notre société, ça, c'est le rôle fondamental du journaliste. Tous les journalistes ne font pas ça. Ils ne sont pas en capacité de faire ça pour le moment. Ils voudraient peut-être bien être les chiens de garde de la démocratie. Moi, je n'en connais pas beaucoup qui le sont, malheureusement. Donc, si ces outils permettent d'abattre les tâches ingrates et de faire en sorte que les journalistes redeviennent des chiens de garde de la démocratie, et bien tant mieux (entretien du 30/03/2020 avec Ricardo Gutiérrez).

Ricardo Gutiérrez poursuivait en expliquant que plutôt que de considérer les générateurs automatiques de textes comme un danger, la FEJ y voyait au contraire une chance pour les journalistes de renouer avec leurs fondamentaux ainsi que leur devoir de base, qui ne consiste pas à « recopier bêtement un communiqué de presse », mais bien à se poser des questions, à réaliser de l'investigation et de l'analyse (entretien du 30/03/2020 avec Ricardo Gutiérrez).

2.5.4. Nouvelles habitudes de consommation

Un autre élément important concernant la survie du journalisme humain concerne l'évolution rapide et permanente des manières de faire du journalisme et de consommer les médias. Par ailleurs, ces comportements de consommation diffèrent selon les groupes d'âge. Il a été démontré que les membres de la génération Y et Z (nées après 1994) préfèrent les articles qui contiennent moins de textes, plus de contenus visuels et présentent plus d'interactivité. Le nouveau mode de récit qui s'est développé en réponse à cette demande implique une intégration de nouvelles plateformes, de nouvelles technologies immersives comme la réalité virtuelle, la réalité augmentée et des applications qui génèrent un engagement dynamique et immersif de la part du consommateur. Les nouvelles méthodes de consommation mettent également l'accent sur l'intégration de clips vidéo (Latar, 2018).

Actuellement, les générateurs automatiques de textes sont limités à la rédaction d'articles routiniers et ne sont pas capables de générer du contenu adapté aux attentes des générations Y et Z. L'IA faible est effectivement incapable de créer un récit journalistique tel qu'espéré par les consommateurs d'aujourd'hui, qui souhaitent une fusion entre des technologies et des applications innovantes. Les algorithmes sont également incapables de s'adapter en permanence aux nouvelles technologies et applications ; leur processus d'élaboration se montre long et n'est pas modulable (Latar, 2018). S'il est un jour possible de créer une IA forte, la situation pourrait changer et les algorithmes deviendraient à même de dépasser les humains dans ces domaines. Cependant, cette éventualité n'est pas annoncée dans un futur

proche (entretien du 9/04/2020 avec Claude de Loupy). Les journalistes, eux, possèdent la capacité de s'adapter aux demandes des nouvelles générations, ce qui représente un autre avantage important par rapport à l'IA (Latar, 2018).

2.5.5. Objectivité

L'idéal d'objectivité trouve ses racines historiques dans le mouvement de naturalisme scientifique du 20^e siècle, qui mettait l'accent sur l'art des méthodes et la reproduction de la réalité avec une objectivité parfaite. Comme expliqué précédemment, dans l'opinion moderne, l'objectivité se révèle un concept problématique car il a été établi qu'il était impossible d'émettre un jugement non subjectif, même inconsciemment. L'idéal d'objectivité des journalistes a ainsi toujours été caractérisé par un paradoxe : lorsque les journalistes tentent de fournir des représentations impartiales du monde, ils deviennent partiaux de par la reproduction de l'ordre social existant. Certains considèrent le journalisme automatisé comme un moyen de contrer la subjectivité inconsciente des journalistes humains. (Milosavljević & Vobič, 2019). Les générateurs automatiques de textes représenteraient ainsi une opportunité d'améliorer la qualité du journalisme actuel (Wu, 2019). L'exemple de *Knowherenews*, abordé dans le chapitre premier, vise notamment cet objectif.

L'argument de l'objectivité des logiciels d'IA suppose évidemment que les données utilisées pour créer un article soient correctes et que les algorithmes soient programmés de façon à rédiger objectivement (Graefe, 2016). Ombelet et al. (2015) soulignent que les générateurs automatiques de textes peuvent également être subjectifs. Les algorithmes de Narrative Science et Automated Insights peuvent effectivement ajuster le ton et la structure d'un article en fonction des profils et des préférences des lecteurs, ce qui s'écarte du modèle d'un article entièrement neutre (Ombelet et al., 2015). En outre, les données utilisées par les algorithmes peuvent elles-mêmes se montrer incomplètes ou tendancieuses, ce qui générerait un texte subjectif (Graefe, 2016). La façon d'écrire des algorithmes peut aussi refléter la subjectivité de leurs concepteurs, consciemment ou non. Les codes des générateurs automatiques de textes sont effectivement créés par des humains qui, eux, ne sont pas objectifs et les algorithmes peuvent donc être influencés par leurs valeurs et leurs partis pris (Milosavljević & Vobič, 2019). Selon Milosavljević et Vobič (2019), cette promesse d'objectivité algorithmique n'est qu'une fiction créée par les développeurs des algorithmes. Il est néanmoins pertinent de mentionner qu'une plateforme inédite, telle que *Knowherenews*, qui

génère plusieurs versions d'un même article (positive, négative, neutre, de droite ou de gauche), fournit aux lecteurs les moyens de se représenter la réalité de façon complète et nuancée et leur permet notamment de ne pas s'enfermer dans la bulle de filtres.

2.5.6. Opportunités pour les entreprises médiatiques

Les opportunités de la génération automatique de textes sont nombreuses pour les entreprises médiatiques (Graefe, 2016 ; van Dalen, 2012), pour qui les articles sont indissociables de la valeur marchande de l'information (Ringoot, 2014). La première opportunité concerne l'efficacité des robots-rédacteurs, qui, contrairement aux journalistes, ne se montrent jamais fatigués ou distraits. Si l'on considère l'hypothèse qu'ils sont programmés correctement et qu'ils utilisent des données fiables, ils ne réalisent pas d'erreurs d'inattention telles que des fautes d'orthographe ou de calcul et ne négligent aucune donnée. Les algorithmes seraient donc moins sujets aux impairs que les journalistes. Lou Ferrara ancienne vice-présidente d'*Associated Press*, a d'ailleurs affirmé que depuis qu'*AP* utilise la génération automatique de textes, le pourcentage d'erreurs au sein de leurs articles concernant le montant des bénéfices des entreprises a chuté de 7 % à 1 % (Graefe, 2016). Il ne faudrait néanmoins pas considérer que les algorithmes ne réalisent aucune faute. Des incohérences au sein des sources peuvent exister, avec des conséquences sur la pertinence de l'article. Cependant, ces incohérences ne proviennent généralement pas du logiciel lui-même, mais bien des données utilisées pour générer l'article (Diakopoulos & Koliska, 2017).

Il existe deux avantages économiques évidents à l'utilisation du journalisme automatisé pour les entreprises. Le premier renvoie à l'augmentation de la vitesse de diffusion des articles (Graefe, 2016). Un texte peut effectivement être mis en ligne une seconde après la réception des données le composant (Raynaud & Didier, 2018). Dans le cadre des matches sportifs, des résultats d'élections ou des rapports de marchés financiers, la rapidité est évidemment d'une importance majeure et permet aux médias de se différencier de la concurrence, ce qui représente une opportunité considérable. En outre, ces textes peuvent par la suite être enrichis par un journaliste qui y apportera de la profondeur et une analyse (entretien du 9/04/2020 avec Claude de Loupy).

Le deuxième avantage économique concerne l'augmentation du nombre d'articles publiés. Le journalisme automatisé permet effectivement de couvrir une plus grande quantité

d'informations en générant des articles qui, à cause des ressources limitées, n'auraient pas pu être rédigés auparavant (Graefe, 2016). Les générateurs automatiques de textes permettent ainsi de compléter une offre d'information insuffisante, parfois même inexistante (Raynaud & Didier, 2018). Dans le cas de *SudPresse*, par exemple, les logiciels de la société LabSense ont donné l'opportunité au média de publier des résumés de matches de football amateur qui n'étaient auparavant pas couverts et que les journalistes n'auraient physiquement pas été capables de rédiger (entretien du 30/03/2020 avec André Thiel). Même si ce genre d'articles n'attire que quelques lecteurs, ils permettent d'augmenter le trafic total du site internet d'une entreprise médiatique. La somme des quelques lecteurs supplémentaires par article représente *in fine* un grand nombre de nouveaux lecteurs qui n'étaient pas adressés auparavant, ce qui a notamment des répercussions positives quant aux revenus publicitaires (Graefe, 2016). Émilie Mascia, rédactrice en chef d'*ELLE à table Belgique* expliquait que, de nos jours, l'augmentation du trafic sur le site internet d'un média représente un élément essentiel :

Chez *ELLE*, notre but, c'est toujours d'augmenter le trafic sur le site internet dans l'objectif final de gagner de l'argent. La presse pour l'instant, c'est difficile. Donc le web c'est un peu le dernier recours où on peut générer de l'argent. Si travailler avec ces données nous permet d'arriver à ce but-là, bien sûr, nous sommes partants (entretien du 20/01/2020 avec Émilie Mascia).

Concernant ces avantages économiques, Flew et al. (2012) rappellent néanmoins l'importance des coûts d'acquisition des logiciels. Plusieurs solutions sont cependant disponibles de façon à rendre le journalisme automatisé plus abordable pour les entreprises, comme des plans de financement alternatifs tels que le *crowd-sourcing* (Thusman, 2019). Pour développer son projet, *Mediafin* a, par exemple, obtenu un financement de 210 000 euros grâce au Digital News Innovation Fund (DNI), soutenu par Google⁷¹.

Le journalisme automatisé représente également un moyen d'améliorer le référencement des articles d'une entreprise médiatique sur les moteurs de recherche (Raynaud & Didier, 2018). Un référencement optimal se révèle aujourd'hui vital pour tout média possédant une plateforme en ligne (entretien du 20/01/2020 avec Marie Guérin, rédactrice en chef d'*ELLE Belgique*). En outre, les articles générés automatiquement peuvent servir de source d'information en cas d'indisponibilité imprévue d'un rédacteur. Si un journaliste ne peut

⁷¹ Source : <https://newsinitiative.withgoogle.com/dnifund/insights/quotebot-editorial-led-robot-delivered-insight/>

exceptionnellement pas se rendre au travail, il pourra être remplacé au dernier moment par les logiciels d'IA. Même si ces derniers se révèlent moins qualitatifs, ils pourront néanmoins être utilisés en tant que remplacement provisoire et venir en aide à un média (entretien du 9/04/2020 avec Claude de Loupy). Du côté des sociétés de génération automatique de textes, le bilan se montre bien évidemment positif. Claude de Loupy affirme :

Tout le monde y gagne : les rédactions peuvent imaginer des projets d'envergure avec la création de dizaines de milliers d'articles, tout en se concentrant sur la valeur éditoriale journalistique, les médias voient leur fréquentation et leur audience augmenter et nous augmentons les possibilités de notre moteur de rédaction (entretien du 9/04/2020 avec Claude de Loupy).

La génération automatique de textes représente d'autres points positifs, précisément pour les médias locaux. Selon Claude de Loupy, les logiciels permettent d'alimenter les lecteurs d'informations « hyperlocales » tout en conservant la rentabilité du média. Grâce au journalisme automatisé, il est effectivement possible de couvrir facilement de nombreux sujets singuliers concernant des petites villes ou localités (entretien du 9/04/2020 avec Claude de Loupy). C'est notamment ce qui s'est passé lorsque le journal *Le Monde* a publié plus de 30 000 articles concernant les résultats de vote de chaque commune française en 2015 (Bronner, 2015). En outre, pour les médias locaux, la couverture « hyperlocale » permet de fidéliser un lectorat qui trouvera les informations qu'il recherche et qui le concerne directement (Raynaud, Didier, 2018).

Par ailleurs, une corrélation a été établie entre l'existence des médias locaux et la préservation des modèles de démocratie en place. En effet, puisqu'elle permet d'exposer les différences qu'il peut exister entre la périphérie et le centre, il a été démontré que la presse locale développe la citoyenneté participative et les valeurs démocratiques (Gürses, 2015). Aux États-Unis, de nombreux médias locaux ont récemment disparu, provoquant une crise de la presse locale. Cette diminution de la quantité de médias locaux, a, selon certains, contribué à amplifier la crise de la démocratie américaine. L'absence d'une presse locale favoriserait en effet la polarisation de la vie politique. Aux endroits où les journaux locaux disparaissent, ou se font plus rares, les citoyens auraient tendance à constamment favoriser un seul et même parti ; au contraire des médias locaux, les médias nationaux privilégient les débats clivants. En outre, en l'absence de médias locaux, les journalistes se trouvent tous concentrés dans les grandes villes, ce qui déforme et altère la perception d'un pays. Les médias locaux se révèlent dès lors indispensables à la démocratie (Rauline, 2019). En les aidant à alimenter leurs pages,

les sociétés de génération automatique de textes permettraient ainsi de préserver l'existence des modèles de démocratie actuels, tout en aidant les médias locaux à traverser la crise de la presse. Il s'agirait, selon Claude de Loupy, « d'une aide parmi d'autres » (entretien du 9/04/2020 avec Claude de Loupy).

2.6. Réflexion

Nous pouvons déduire de cette analyse que les conséquences de l'utilisation de la génération automatique de texte en presse écrite posent un certain nombre de questions et se caractérisent par un grand niveau d'incertitude. D'une part, cette nouvelle technologie apporte de nombreuses opportunités aux entreprises médiatiques et permettrait non seulement aux journalistes de retrouver leur rôle de chien de garde de la démocratie, mais aussi de récupérer du temps pour apporter une réelle plus-value à leurs productions. D'autre part, le journalisme automatisé représente une menace importante pour ces mêmes journalistes s'il venait à les remplacer. Cette menace dépend notamment du niveau de contingence concernant les capacités des algorithmes à égaler, voire un jour à dépasser les capacités créatives du cerveau humain. Elle concerne également l'incertitude concernant l'évolution du NLP et du NLG, à savoir si les algorithmes seront, un jour, capables de comprendre toute la richesse et la profondeur du langage, sans cesse en évolution, et à traduire cette richesse en des textes lisibles. Il s'agit donc ici du débat concernant l'existence éventuelle d'une IA forte, comme prédite par les pères fondateurs de l'IA et plus récemment par Kahneman et Kurzweil. Cette école de pensée est cependant réfutée par Searle, Boden, ou encore Andreassen, selon lesquels les capacités créatives de l'IA seront toujours inférieures à celles du cerveau humain. Ce point de vue présente un atout important pour les journalistes. Les incertitudes concernant l'existence d'une IA forte témoignent néanmoins du besoin fondamental d'élaborer un cadre juridique, éthique et déontologique clair quant à l'utilisation des générateurs automatiques de textes en presse écrite, notamment de façon à préserver l'intégrité du métier de journaliste et à garantir son rôle de quatrième pouvoir au sein de nos sociétés démocratiques dans le cas où l'IA viendrait à dépasser l'ensemble des capacités humaines.

CHAPITRE 3

Étude de contenu

3.1. Présentation de l'étude de contenu

Après avoir détaillé le développement de la génération automatique de textes en presse écrite et avoir relevé les questions qu'elle pose d'un point de vue légal, éthique et sociétal, il est pertinent de s'intéresser *in fine* à la nature des articles générés automatiquement. Ce troisième chapitre comporte dès lors une courte étude de contenu d'une sélection d'articles. Notre choix s'est porté sur les résumés de matches de football amateur générés par la société LabSense sur le site internet de *SudPresse*. Nous nous intéresserons ainsi à la seule société de presse écrite belge qui génère à ce jour des articles automatiquement. Cette étude ne se veut aucunement représentative de ce qu'est capable de produire un logiciel de génération automatique de textes. Il s'agit ici, à petite échelle, d'illustrer une des manières dont cette technologie est utilisée en journalisme en Belgique. Dans ce cadre, nous envisagerons uniquement le contenu des articles sélectionnés et non pas leur réception, cet aspect ayant déjà été fréquemment analysé dans d'autres études, comme précisé dans l'introduction. Au vu de la simplicité des articles, cette analyse se voudra brève, mais permettra au lecteur de se représenter plus précisément la forme et le contenu de textes rédigés et publiés sans intervention humaine en presse écrite.

La sélection de textes analysés dans ce chapitre est composée de trente articles du site de *La Meuse*⁷². Il s'agit de résumés de matches de football amateur en province de Liège (matches provinciaux et régionaux hommes et femmes). Les articles ont été sélectionnés de façon aléatoire sur une période s'étendant de décembre 2019 à février 2020. Ces résumés seront comparés à d'autres articles de football rédigés par des journalistes et publiés sur le site de *La Meuse* sur la même période de temps. Ces articles sont au nombre de quinze et couvrent des matches de la Jupiler Pro League, de La ligue des champions, de la Coupe de Belgique et de football étranger. Les liens des articles sélectionnés se trouvent dans les annexes de ce travail⁷³. Pour réaliser cette étude, nous nous sommes inspirés du livre *Analyser le discours de presse*, de Roselyne Ringoot (2014) et du tableau d'analyse qu'elle propose. Nous avons

⁷² Source : https://www.lameuse.be/41951/sections/resultats-foot-regional-liege?fbclid=IwAR2BmLwm_wU-b8VizvT37jhNEGM43iw9T2JzsRLiFyyj5uwCygczvK0QyDg

⁷³ Voir annexes 10, p.142.

également utilisé RapidTables⁷⁴, une plateforme permettant de calculer la fréquence des mots dans un texte (jusqu'à trois mots), de façon à déterminer si des expressions ou des tournures de phrases reviennent fréquemment au sein des articles rédigés automatiquement.

3.2. Observations générales et cadrage

Les résumés de matches de football amateur générés automatiquement sur le site de *La Meuse* font partie de la rubrique « sports » du site internet principal. On peut également y accéder via l'adresse monclubdefoot.be en encodant le nom d'un club de foot, d'un comité ou la compétition recherchée. Les articles apparaissent alors sur une page intitulée « Résultats foot régional Liège » d'une dominance de couleur orange, en accord avec la section sport du site. Ces articles sont relativement courts. Sans compter la feuille de match, les trente articles varient entre 788 et 1807 signes (espaces compris) et présentent une moyenne de 1249 signes (espaces compris). Cette longueur se rapproche davantage du format des dépêches d'agences de presse que des articles de la section sport d'un journal.

Les titres de l'ensemble des articles sont très similaires les uns des autres au niveau de la structure et des informations qu'ils livrent. Ces derniers sont de nature informative, en opposition aux titres incitatifs, selon la classification de Roselyne Ringoot⁷⁵ (2014). On retrouve des intitulés tels que : « Provinciale III série C (Liège) : succès 2-3 du Royal Aubel Football Club face au S.C. Saive », « U14 Provinciaux série B (Liège) : succès 5-2 du Football Club Herstal face au R.E.S. Templiers-Nandrin », ou encore « Provinciale I Dames (Liège) : égalité 2-2 entre le S.C. Saive et le Royal Association Football Franchimontois ». Ces titres sont relativement longs et ont systématiquement recours au double point, typique des titres informatifs, qui organise le rapport entre le thème et le rhème des deux parties de l'intitulé (Ringoot, 2014). Pour chaque titre, le thème (précédant le double point) indique la compétition au cours de laquelle le match résumé dans l'article a eu lieu. Le rhème (succédant le double point), lui, donne le score du match et les noms des deux équipes qui se sont affrontées. On constate ainsi l'existence d'un titre-modèle, où les informations récoltées sur

⁷⁴ <https://www.rapidtables.com/text/word-frequency.html>

⁷⁵ Roselyne Ringoot (2014) identifie deux types de titres en presse écrite. Elle explique : « le discours professionnel distingue deux “styles” de titres qui sont les titres informatifs et les titres incitatifs, renvoyant au “style du *Monde*” et au “style *Libé*”. Ces deux modèles à partir desquels s'effectuent des choix énonciatifs plus ou moins nuancés opposent deux rapports à l'information. Le premier promeut la neutralité et l'objectivité en répondant en partie aux impératifs de la règle des 5W (quoi, où, quand, pourquoi, comment). Le second privilégie l'effet de surprise visant à susciter la curiosité du lecteur » (p. 89).

la feuille de match sont utilisées pour former un titre cohérent. Ces intitulés sont neutres et répondent uniquement à un objectif informatif. Ils ne comportent effectivement aucun effet de surprise, pas de jeux de mots, d'effet de style ou de recherche. On ne distingue pas d'envie de susciter la curiosité du lecteur, mais uniquement de l'informer en répondant aux questions « qui ? », « quoi ? » et « comment ? ». Comme l'indique Ringoot (2014), les titres adoptent la fonction de guide⁷⁶ en annonçant et définissant quelle est l'information principale d'un article. Dans ce cas-ci, il s'agit du résultat du match et du déroulement de celui-ci, uniquement. Ces titres ne peuvent pas être qualifiés d'accrocheurs et ne répondent pas aux habitudes des médias sur internet à rendre leurs titres racoleurs. On parle régulièrement de « pièges à clics » (« clickbait » en Anglais), ou, plus vulgairement, de « putaclics » (Falcinelli, 2019). Les titres des articles générés automatiquement sur le site de *La Meuse* ne s'inscrivent aucunement dans cette tendance, présentant des intitulés longs qui dévoilent directement l'information principale de l'article.



Figure 11. Exemple de la photo accompagnant les résumés de matches de football générés automatiquement sur le site de *Sudpresse*.

Sous le titre, une image, toujours identique, reprend les informations importantes du match : la compétition, le numéro de la journée, la date, le nom de l'arbitre, le score et les noms des deux équipes. Il s'agit d'une image de terrain de football vide sur laquelle les indications textuelles sont ajoutées. Cette photo souligne une nouvelle fois que l'information principale de l'article est le résultat du match et son déroulement. Cette image n'est pas générée par LabSense, mais bien en collaboration avec la société IDalgo, spécialisée dans les

⁷⁶ Ringoot souligne : « Les titres d'information sont d'une importance capitale ; car non seulement ils annoncent la nouvelle (« la fonction épiphanique »), non seulement ils conduisent à l'article (« fonction guide »), mais encore, ils résument, ils condensent, voire ils figent la nouvelle au point de devenir l'essentiel de l'information » (p.87).

infographies sportives (entretien du 30/03/2020 avec André Thiel, directeur *Process & Development* chez *SudPresse*). Sous l'article, on retrouve deux autres infographies, l'une résumant les résultats du jour dans la compétition concernée et l'autre reprenant le classement des différents clubs. Cependant, l'article n'est accompagné d'aucune photo représentant le déroulement du match, ces dernières ayant pourtant été systématisées dans la presse d'information générale (Ringoot, 2014). Cela s'explique bien évidemment par l'absence d'un journaliste ou d'un photographe sur place, toutes les données étant récoltées via l'URFBSA. Ces informations permettent néanmoins de créer des éléments visuels qui illustrent l'article.

Les chapeaux des résumés de matches, de façon similaire au schéma selon lesquels les titres sont créés, livrent tous les mêmes informations, à savoir, le score, le nom de l'équipe gagnante et perdante, la journée de match et le lieu où il s'est disputé. Le chapeau ajoute dès lors les questions « quand ? » et « où ? » aux renseignements déjà fournis par le titre. Nous pouvons donner l'exemple type de chapeau suivant : « Sur un score de 8 à 2, la Royale Alliance Melen-Micheroux a remporté le match de la 17ème journée l'opposant au Royal Herve Football Club à domicile... ». Ainsi, les chapeaux font généralement de deux à trois lignes et se terminent systématiquement par des points de suspension. Quant à la structure, certains chapeaux commencent par indiquer le résultat du score alors que d'autres commencent par donner le numéro de la journée de match, par exemple. On note également des alternances entre l'utilisation des mots « a remporté » et « a gagné ». En cas d'égalité, la phrase est tournée légèrement différemment. On retrouve notamment les expressions « ne sont pas parvenus à se départager », « match nul » et « égalité » qui sont utilisées à tour de rôle. En outre, en fonction de l'endroit où s'est disputé le match, la préposition « au » ou « sur » est utilisée. Si les chapeaux livrent tous les mêmes informations, on note néanmoins quelques différences dans l'emploi des mots et la structure des phrases, contrairement aux titres qui sont tous similaires.

Sous le chapeau de chaque article, est indiqué : « Note importante : cet article a été rédigé sur base des données officielles de l'URBSFA ». *SudPresse* applique donc une politique de transparence totale quant à l'origine des sources. Cependant, aucun élément ne mentionne que l'article a été écrit par un robot-rédacteur, ce qui va à l'encontre de l'avis de nombreux journalistes et chercheurs, selon lesquels cette information devrait être divulguée au lecteur, comme précisé dans le deuxième chapitre de ce travail. Dans ce cas-ci, pourtant, un lecteur

non averti pourrait juger qu'il s'agit ici d'articles rédigés par des journalistes, d'autant que *SudPresse* ne mentionne aucun partenariat avec LabSense sur son site internet.

Par ailleurs, il est pertinent de mentionner que les résumés de matches de football amateur ne semblent pas être générés dans l'optique d'attirer l'attention de la génération Z, qui préfère notamment les vidéos et les plateformes plus interactives (Latar, 2018). Ces articles s'avèrent destinés prioritairement aux parents des enfants jouant en club de football. Cela se ressent dans la présentation des informations et c'est en outre à eux que s'adresse la page d'accueil « monclubdefoot.be », qui donne accès à tous les résumés de matches de football amateur :

Eden Hazard, Kevin De Bruyne, Thibaut Courtois... Ces idoles ont toutes foulé les pelouses de Braine-leComte et d'ailleurs, là où se forment les futures stars de la Ligue des Champions. Vous êtes d'ailleurs des centaines de milliers, chaque week-end, à vous y rendre avec vos enfants⁷⁷.

Le public ciblé par *SudPresse* se révèle donc clairement être les parents d'enfants pratiquant le football amateur en club.

3.3. Hiérarchisation des informations

Tout comme les éléments de cadrage qui sont identiques au sein de chaque article, la hiérarchisation des informations du corps du texte suit, elle aussi, un modèle-type. Les résumés de matches adoptent la technique de la pyramide inversée, qui vise à livrer l'essentiel des informations en début d'article et à les préciser au fil des paragraphes (Ringoot, 2014). Le premier paragraphe de l'article reprend tous les éléments évoqués dans le titre et le chapeau et les reformule en un texte continu. Certaines informations sont parfois ajoutées, telles que les noms des joueurs qui ont dominé le match, mais cela ne représente pas la majorité des articles. Le deuxième paragraphe, lui, consiste en un bref résumé du match, qui indique notamment à quelle minute ont eu lieu les attaques décisives, les buts, le score à la mi-temps et le résultat final. Il est également mentionné par qui les buts ont été marqués et s'ils ont été réalisés par penalty. On notera notamment de nombreuses nuances dans la façon dont ce paragraphe particulier est rédigé, concernant à la fois à la structure et les termes utilisés. Le troisième paragraphe, lui, indique si la confrontation a été riche ou exempte de fautes et à qui l'arbitre a adressé des avertissements.

⁷⁷ Source : <http://monclubdefoot.be/#how-it-works>

Sur un score de 4 à 0, la Royale All. Clavinoise S.C. a remporté le match de la 18ème journée l'opposant au Grupo Desportivo Poulseur à domicile...

Note importante : cet article a été rédigé sur base des données officielles de l'URBSFA

La [Royale All. Clavinoise S.C.](#) a accueilli le [Grupo Desportivo Poulseur](#) au Complexe rue Du Marché pour un [match de Provinciale IV série B \(Liège\)](#). Ce match était arbitré par Damien Duhem. Victoire 4 à 0 pour la Royale All. Clavinoise S.C. dans cette rencontre face au Grupo Desportivo Poulseur. Les supporters ont assisté à un match riche en buts, grâce à Regis Becker et Cyril Pollain.

Les Claviéris ont pris les commandes dans cette confrontation en ouvrant la marque à la 36ème, grâce à Regis Becker. Les Condruziens ont vu leur avance au score se creuser, grâce à Cyril Pollain à la 41ème. À la fin de la première mi-temps, les joueurs sont rentrés aux vestiaires sur un score de 2-0. Avec le but de Regis Becker à la 86ème, la Royale All. Clavinoise S.C. a alourdi la facture. Avec son but à la 91ème, dans le temps complémentaire, Cyril Pollain a corsé l'addition.

L'arbitre a sorti une carte pour Jean-Claude Aelens, puis pour Antoine Rosimont, Antoine Imhauser et Loïc Yves Aelens.

Figure 12. Exemple de la structure des trois premiers paragraphes d'un article rédigé automatiquement sur le site de La Meuse⁷⁸.

Ces premiers paragraphes sont systématiquement suivis d'un sous-titre indiquant les noms des deux équipes ayant disputé le match, accompagné de la mention « : les prochains matches ». Un court paragraphe indique alors les futures rencontres de chacune des deux équipes. Ce paragraphe est suivi d'un dernier sous-titre intitulé « La feuille de match » pour les matches provinciaux et « La rencontre en bref » pour les matches régionaux. Cette section n'est pas rédigée de manière continue. Sont alors indiqués la composition de chaque équipe, le nom de l'arbitre et éventuellement les buts marqués.

Même si tous les articles suivent cette hiérarchie, certains d'entre eux sont occasionnellement beaucoup plus courts que les autres. Le paragraphe de la description du match se résume effectivement parfois à une phrase ou n'est tout simplement pas présent. On s'aperçoit ici de la dépendance directe du générateur automatique de textes sur les données du match, sans lesquelles il est incapable de rédiger. Les articles sont donc plus courts en fonction des données qui ont été encodées et dont le logiciel dispose. Aucun des matches régionaux de notre sélection d'articles ne présentait un paragraphe résumant le déroulement du match, par exemple. Même au sein des articles provinciaux, on s'aperçoit que la base de données détermine étroitement la longueur de l'article, qui varie sensiblement d'une rencontre à l'autre.

⁷⁸ Source : <https://www.lameuse.be/490273/article/2019-12-15/provinciale-iv-serie-b-liege-succes-4-0-de-la-royale-all-clavinoise-sc-face-au>

3.4. Rhétorique et thèmes abordés

Selon la classification de Ringoot (2014), les articles automatisés de notre sélection se rapprochent du genre de la neutralité⁷⁹. On vise ici à une dépersonnalisation, l'information parle d'elle-même ; il s'agit d'une information brute qui échapperait à toute subjectivité. Les articles adressent « la règle des 5W » du jargon professionnel du journaliste : « *who* ? », « *what* ? », « *when* ? », « *where* ? », « *why* ? », et ce, dès le chapeau, ce qui correspond aux règles qui se sont généralisées au 20^e siècle dans les écoles de journalisme et les agences de presse (Bastin, 2018). Ringoot (2014) explique que ces cinq questions sont toujours complétées par un sixième « *how* ? ». Cependant, cette dernière question n'est pas adressée dans tous les articles puisque certains d'entre eux sont exempts de paragraphe résumant les principales attaques du match ou que ce dernier se résume parfois à une seule phrase en fonction des données à disposition du logiciel de génération automatique de textes.

L'angle sous lequel les articles sont traités est celui de la performance. L'événement⁸⁰ (Ringoot, 2014) est le match en lui-même et ne s'étend pas à d'autres éléments pouvant s'être produits avant ou après le match. Les sujets de l'article sont systématiquement les joueurs et l'arbitre. On ne constate, par exemple, aucune indication concernant des conditions de jeu particulières ou encore le nombre de supporters et leur implication dans le jeu, cela n'étant pas repris dans les feuilles de match de l'URBFSA. En outre, le point de vue est extérieur, l'état d'esprit ou le ressenti des joueurs n'étant pas décrits. On ressent fortement que les articles sont écrits grâce à une base de données et non pas à partir d'interviews ou d'une présence physique. Concernant le code de déontologie journalistique, rien ne fait l'objet de transgression des règles. D'un point de vue juridique, nous ne constatons aucun élément qui pourrait porter atteinte à l'intégrité d'un individu ou poser question.

Le langage utilisé est formel. On ne constate effectivement aucune interpellation du lecteur ou de formule familière ou humoristique. Mis à part les sections « La feuille de match » ou « La rencontre en bref », qui ne sont pas écrites de manière continue, les autres paragraphes des articles s'enchaînent de manière correcte et naturelle. Par ailleurs, même sans avoir

⁷⁹ Ringoot (2014) identifie quatre genres journalistiques principaux : le vécu, la révélation, le point de vue et la neutralité.

⁸⁰ Selon Ringoot (2014), « En journalisme, l'événement s'inscrit à la fois au cœur des pratiques professionnelles et au cœur de la dimension éditoriale. Le journal contribue à « construire » l'événement en attribuant une valeur particulière à certaines informations retenues selon sa périodicité » (p. 78).

assisté au match et uniquement grâce aux données reçues, le logiciel est capable de livrer des informations sur le déroulement du match. Il indique effectivement qu'une victoire a été « sans appel et sans suspense », « facile », ou encore qu'une rencontre s'est montrée « serrée ». Il utilise également des formulations fréquemment employées par les journalistes sportifs, telles que « en faveur des rouge et blanc », « les bleus », « À la pause, chaque équipe a repris le chemin des vestiaires », « un match riche en buts », « Il a rapidement forcé le gardien du Royal Aubel Football Club à aller chercher la balle dans ses filets », « Andrea Wauthier a remis les compteurs à zéro en égalisant », « Nicola D'Iuliiis a trouvé la faille ». Ce genre de formulations donne un côté plus humain à l'écriture. On constate même l'utilisation de métaphores dans des phrases telles que : « Simon Pelsser a éclaboussé le match de son talent », « Vincenzo Giuliano a permis à son équipe de prendre le large », « le contrat est rempli », « hécatombe sur le terrain » ou encore « le scénario a continué de sourire ». Ce type de métaphores ne domine pas l'ensemble des articles, mais on peut néanmoins affirmer qu'elles permettent à l'ensemble des résumés de matches d'adopter un style d'écriture plus naturel.

En outre, nous n'avons relevé aucune faute d'orthographe dans les articles rédigés automatiquement, ce qui est en accord avec les affirmations faites dans le deuxième chapitre. Cependant, nous avons relevé une faute grammaticale impliquant un déterminant dans la phrase : « Pour son prochain match, le Racing Club Vaux-Chaufontaine affrontera l'U.S. Grace-Hollogne (7^{ème} du championnat) à Piscine Communale le 8 décembre ». Les règles de grammaire veulent que le déterminant « la » précède le mot « piscine ». Ce genre de faute est revenu à deux reprises. On ressent ici encore une fois l'utilisation de phrases modèles où des noms ont été ajoutés dans des cases. En outre, certaines expressions témoignent parfois d'une certaine lourdeur, telles que : « Royale-etoile-sp-dalhemoise sera face au S.C. Saive le même jour lors de la même journée, au stade F. Fassotte ». Si cette phrase est compréhensible, un journaliste éviterait généralement les répétitions des mots « même » et du préfixe « jour ». De manière générale, on remarque plusieurs répétitions de mots au sein d'un même article.

3.5. Fréquence des mots

À première vue, les articles paraissent très semblables les uns des autres. Mis à part la hiérarchisation des informations, qui se révèle identique d'articles en articles, les mots et les formulations utilisées semblent également revenir à plusieurs reprises. Pour déterminer à quel

point ces articles sont répétitifs, nous avons utilisé la plateforme RapidTables, permettant de calculer combien de fois un groupe de mots se présente au sein d'un texte, jusqu'à trois mots. Nous avons couplé cette technique à une simple recherche « Ctrl f » dans notre corpus de textes, pour élargir les groupes à plus de trois mots si nécessaire. Nous n'avons bien évidemment pas tenu compte des noms d'équipes de football et des lieux de matches.

Les groupes de mots qui sont revenus le plus souvent sont : « un score de » (26), « sur un score » (14), « un match de » (14) suivi du nom de la catégorie de la compétition, « gagné le match » (13) , « remporté le match » (15), « pour son prochain match » (16), « exactement au même moment » (12), « sera face » (14), « le match contre » (12), « pour la [chiffre]ème journée » (12), « écopé d'une carte jaune » (9), ultime ou dernier « but de la rencontre » (12), « fin de la première » (7) suivi des mots « période » ou « mi-temps », et « a trouvé la faille » (4). Ces résultats montrent que plusieurs groupes de mots reviennent effectivement à de nombreuses reprises dans les articles sélectionnés et indiquent un certain niveau de répétition. Cependant, il est pertinent de noter qu'aucun groupe de mots n'est revenu trente fois, ce qui signifie que le logiciel tente de varier les tournures de phrases dans la mesure du possible et n'utilise pas exactement les mêmes formulations pour chaque article. En outre, on s'aperçoit que les groupes de mots répétitifs sont présents de manière importante au sein du chapeau, du premier paragraphe, ainsi que du paragraphe rappelant les prochains matches de chaque équipe. La section détaillant le déroulement du match (deuxième paragraphe), elle, présente moins de groupes de mots fréquents et semble donc moins répétitive. Ce paragraphe arbore effectivement plus de variétés au niveau de l'emploi des mots et des tournures de phrases. Il s'agit notamment d'un point sur lequel *SudPresse* porte de l'importance. André Thiel expliquait effectivement qu'il cherche aujourd'hui à poursuivre le développement de leur logiciel de manière à créer des articles moins similaires les uns des autres (entretien du 30/03/2020 avec André Thiel).

3.6. Comparaison entre articles automatisés et articles traditionnels

De façon à nuancer notre analyse, il est intéressant de comparer les articles rédigés automatiquement avec d'autres articles trouvés dans la section « foot » du site internet de *La Meuse*. La première observation révèle que très peu de textes rédigés par un journaliste ne se cantonnent au résumé d'un match de football. Ces derniers impliquent effectivement très souvent de l'analyse. Les journalistes cherchent fréquemment à comprendre ce qui a fait

défaut à une équipe et à analyser les actions importantes d'une rencontre. Ils expliquent aussi les conséquences de l'absence d'un joueur sur le déroulement d'un match ou encore ce que le résultat d'une rencontre aura comme conséquence pour la suite d'une compétition. En outre, plutôt que de se trouver au centre de l'article, les données et les chiffres sont souvent utilisés pour illustrer les points que les journalistes veulent démontrer. Les données sont d'ailleurs parfois même mises entre parenthèses, comme si celles-ci étaient superflues par rapport à l'analyse, au contraire des articles automatisés où les données prennent la place principale. De façon générale, les articles ne se limitent pas aux événements des matches. Les journalistes y abordent également les transferts ou les blessures de certains joueurs, de nouvelles techniques mises en place, ou encore des amendes qui auraient été infligées. Les journalistes évoquent également des circonstances et des détails extérieurs au match, tels que la météo, l'état du terrain, les discussions dans les vestiaires ou l'ambiance au sein des supporters. Par ailleurs, les journalistes agrémentent aussi leurs articles d'interviews des entraîneurs et des joueurs et intègrent donc plusieurs points de vue à leurs productions. Enfin, sur le site de *La Meuse*, et en dehors des articles sélectionnés pour cette analyse, on constate l'omniprésence de faits divers : un vestiaire retrouvé très sale après un match, l'histoire de la perte de poids d'un joueur, des disputes au sein d'un club, un joueur qui aurait oublié de payer la pension alimentaire de son fils, etc. Dans le domaine du football amateur, il s'agit d'informations dont le logiciel de génération automatique de *SudPresse* n'est pas en possession et qu'il ne saurait donc pas traiter.

Les articles rédigés par les journalistes sont en moyenne significativement plus longs que ceux rédigés automatiquement. La structure est à chaque fois différente, au contraire des articles de football amateur, qui suivent un modèle bien précis. Dans le cas des articles traditionnels, les chapeaux apportent des angles variés et le choix des sous-titres est donné à l'appréciation du journaliste. Concernant les titres, selon la classification de Ringoot (2014), plusieurs d'entre eux sont de nature incitative, tels que « Ligue des champions : L'Athlético Madrid a retrouvé son âme », par exemple, contrairement aux titres des articles automatisés qui étaient systématiquement informatifs. Les intitulés des articles rédigés par les journalistes ne se résument ainsi pas seulement à citer le nom d'un match. Ces derniers sont plus recherchés et comportent souvent des citations, comme par exemple « Meunier : Nous avons un peu manqué de volonté ». On retrouve aussi systématiquement la présence de photos représentant le propos de l'article. De façon générale, les articles sont rédigés dans un style d'écriture plus abouti, sont plus vivants et présentent plus de métaphores, telles que : « a pris

une gifle », « a asphyxié » telle équipe, « a retrouvé son âme », « régné en maître », « écrase », « se quittent dos à dos » ou « sur un nuage ».

3.7. Conclusion de l'analyse

Comme nous pouvions nous y attendre au vu des éléments développés dans les deux chapitres précédents, les résumés de matches de football amateur rédigés automatiquement sur le site de *La Meuse* se distinguent de par leur caractère répétitif et technique. On discerne une structure identique à tous les articles. Les titres et les chapeaux apparaissent comme des textes à trous dans lesquels les informations sont remplies selon les données récupérées au cours du match. Le corps du texte, lui, présente des formulations de phrases et des termes qui diffèrent d'articles en articles. Cependant, seul le deuxième paragraphe apporte de réelles différences concernant la structure, les termes et les expressions utilisées. Par ailleurs, l'angle abordé se révèle toujours identique, les articles évoquent uniquement le déroulement du match. Enfin, mis à part deux fautes de grammaire et quelques lourdeurs, nous avons constaté une utilisation correcte de la langue française et l'utilisation de métaphores. Ces dernières sont néanmoins plus présentes dans les articles rédigés par des journalistes.

L'absence de « plus-value journalistique », comme abordée par les journalistes au cours des entretiens qui ont ponctué ce mémoire, se ressent ici fortement. Selon les journalistes interrogés, cette plus-value se traduit notamment par l'analyse, l'émotion et la pratique des interviews. Ces éléments ne se retrouvent pas dans les articles rédigés automatiquement sur le site de *La Meuse*, mais ils sont bien présents dans les articles rédigés par des journalistes. Cet aspect confirme le fait que les journalistes possèdent actuellement des qualités que les générateurs automatiques de textes ne sont pas encore capables d'atteindre, ce qui représente une opportunité pour ces derniers. Néanmoins, il est important de préciser que les productions analysées dans le cadre de cette courte étude sont basiques et que d'autres articles générés automatiquement, dans d'autres pays et par d'autres entreprises médiatiques, pourraient comporter des différences majeures. Nous avons cependant trouvé pertinent d'illustrer la manière dont cette technologie est utilisée en presse écrite en Belgique.

CONCLUSION

Le journalisme automatisé a été anticipé dès 1965 par Michael Frayn dans sa nouvelle satirique *The Tin Men* (Frayn, 1965). S'il aura fallu plusieurs décennies pour que l'utopie de Frayn devienne réalité, les générateurs automatiques de textes sont désormais utilisés par plusieurs entreprises médiatiques à travers le monde (Dörr, 2016 ; Graefe, 2016). Ce mémoire avait pour première ambition de dresser un historique problématisé de l'utilisation des logiciels de génération automatique de textes en presse écrite. Treize sociétés proposent aujourd'hui leurs services d'automatisation, notamment aux entreprises de presse. Automated Insights et Narrative Science, les deux sociétés pionnières, basées aux États-Unis, ont popularisé et répandu l'utilisation de leurs systèmes dès le début des années 2010. Ainsi, des sociétés telles qu'*Associated Press* et *Forbes* utilisent aujourd'hui abondamment leurs services. D'autres entreprises américaines, telles que le *LA Times* et le *Washington Post* ont choisi de développer leurs propres algorithmes en interne de façon à conserver une indépendance totale. À ce jour, le journalisme automatisé a également conquis d'autres continents. En 2015, *Le Monde* surprenait l'univers du journalisme avec une première mondiale et un travail d'une énorme ampleur : la publication de plus de 30 000 articles générés automatiquement, en une nuit, en langue française.

Depuis peu, la Belgique, qui était à la traîne en la matière, a rejoint le mouvement. Ce mémoire a mis en lumière les productions belges d'articles automatisés de Bxl'air bot et de *SudPresse*, mais aussi le futur projet d'automatisation d'articles financiers de *Mediafin*. *SudPresse* a notamment affirmé vouloir étoffer son offre d'articles générés automatiquement dans les domaines de la météo et travailler à l'amélioration de la qualité de ses articles automatisés. *Mediafin* a également évoqué la possibilité d'étendre l'ampleur de son projet dans le cas où les résultats seraient positifs. Il est pertinent de souligner que la plupart des logiciels mentionnés au cours de ce travail n'ont pas été conçus dans l'optique de remplacer le travail des journalistes humains. La génération automatique de textes y a effectivement été abordée comme une aide à l'écriture, comme une possibilité de libérer du temps pour les journalistes, ou encore comme une solution pour créer du contenu ne pouvant pas être physiquement couvert par les journalistes. Seule la plateforme *Knowherenews*, dans un souci d'objectivité, a désiré promouvoir l'utilisation presque absolue des algorithmes d'IA.

Il est très probable que les technologies de génération automatique de textes se perfectionnent dans les années à venir. Comme nous l'a expliqué Claude de Loupy (CEO de la société Syllabs) dans l'entretien qu'il nous a accordé, les entreprises spécialisées cherchent continuellement à améliorer les logiciels déjà mis en place, et notamment les capacités linguistiques de ces derniers. Les médias, eux, sont de plus en plus intéressés par les services proposés par ces sociétés et commencent à les considérer comme un enjeu de survie à moyen terme (entretien du 9 avril 2020 avec Claude de Loupy). Selon Noam Lemelshtrich Latar (2018), les innovations de pointe en IA permettront bientôt une automatisation entière des *news room*, dans lesquelles les robots auront remplacé les fonctions humaines principales : les éditeurs, les journalistes et le personnel commercial. À ce jour, dans le domaine journalistique, les capacités des logiciels de génération automatique de textes sont déjà nombreuses. Ces derniers sont effectivement capables d'analyser d'immenses quantités de données à une vitesse très importante. Ils peuvent créer des articles sans aucune intervention humaine, suivant l'angle, le ton et la langue choisie par une entreprise médiatique, et en très grande quantité. Nous avons identifié les opportunités que ces logiciels représentent pour les entreprises médiatiques et plus particulièrement pour la presse locale. Ils comportent notamment la quantité et la vitesse de diffusion, permettant de rivaliser avec d'autres plateformes, l'enrichissement des contenus, ainsi que l'amélioration du référencement du site internet de l'entreprise sur les moteurs de recherche. Ces logiciels constituent également une opportunité intéressante pour les journalistes de se livrer à des tâches plus valorisantes que celles imposées par le journalisme routinier et répétitif. Il pourrait effectivement s'agir, pour eux, d'une occasion de réaliser du journalisme de temps long, de se consacrer à l'investigation, à la poursuite d'interviews et aux formats créatifs. Nous avons également abordé la possibilité pour les journalistes de retrouver leur rôle de chien de garde de la démocratie, ou du quatrième pouvoir, qu'ils ont parfois aujourd'hui l'impression de négliger.

Une des inquiétudes significatives du développement de l'IA, cependant, consiste en la disparition du métier de journaliste humain au profit des robots et de l'automatisation. Nous avons établi que cette menace se caractérise par un certain niveau d'incertitude. Des scientifiques éminents, tels que Ray Kurzweil et Daniel Kahneman, défendent l'existence d'une IA forte et affirment que l'IA sera, un jour, capable de surpasser les capacités du cerveau humain (Latar, 2018). D'autres chercheurs, néanmoins, tels que John Searle et Margaret Boden, rejettent cette idée que l'IA pourrait un jour atteindre les capacités du cerveau humain, et particulièrement concernant la créativité. Dans ce cas, il s'agirait d'une

bonne nouvelle pour les journalistes qui conserveraient des avantages majeurs par rapport aux machines, ce qui garantirait leur place dans l'industrie. En dehors de ce débat concernant l'existence d'une IA forte, les journalistes interrogés dans le cadre de ce mémoire, mêmes s'ils sont conscients des avantages notables que pourrait leur offrir la génération automatique de textes, ont émis des craintes sur le fait que les entreprises médiatiques ne finissent par embaucher moins de journalistes au profit des logiciels d'IA.

Ce travail a souligné d'autres enjeux liés à l'utilisation de la génération automatique de textes en presse écrite. Ces derniers impliquent notamment des problèmes de déontologie journalistique et des aspects tels que la transparence ou l'origine des sources. De plus, puisqu'il n'existe encore aucune loi concernant l'utilisation de l'IA en Europe, nous avons identifié des problèmes concernant la responsabilité juridique dans le cas où les articles générés automatiquement contiendraient des propos inappropriés ou offensants. Enfin, les journalistes interrogés dans le cadre de ce mémoire ont fréquemment évoqué la menace posée par les algorithmes d'IA pour la qualité du journalisme de presse écrite, les logiciels n'étant pas capables de ressentir d'émotions, de développer un véritable style d'écriture, d'être subjectifs, ou encore de réaliser un recoupement des sources élaboré.

Les limites actuelles des générateurs automatiques de textes ont ainsi également permis d'identifier des opportunités pour les journalistes de rester en lice sur le marché de la presse écrite. Contrairement aux logiciels d'IA, les journalistes peuvent notamment rédiger des textes à partir de données non quantifiables, avoir recours à des émotions, à des valeurs et à des informations de contexte. Ils sont également capables de produire des interviews, des analyses ou des cartes blanches et ont des capacités d'écriture plus évoluées que les algorithmes, ce que nous avons qualifié de « plus-value journalistique ». L'étude du troisième chapitre a permis de mieux se représenter la structure et le contenu d'un article généré automatiquement, mais aussi de mettre en avant cette plus-value journalistique humaine, qui n'était pas présente au sein des articles générés automatiquement par LabSense pour *SudPresse*. Une autre opportunité pour les journalistes a été identifiée grâce à la nature sans cesse changeante des modèles d'écriture journalistique. Le journalisme est devenu un mélange de nouvelles technologies, incluant la réalité virtuelle, des vidéos à 360°, ou encore des applications telles que Snapchat ou Vine. Les nouveaux contenus contiennent moins de texte et impliquent une interaction importante avec le consommateur, ce qui n'est actuellement pas réalisable par les logiciels d'IA et laisse la place aux journalistes humains.

Même si les algorithmes représentent une menace importante, leurs limites inhérentes offrent ainsi aux journalistes des opportunités de proposer des contenus différents.

Le théoricien de renommée mondiale Stephen Hawking a récemment affirmé que l'IA pouvait être considérée comme la meilleure des inventions dans l'histoire de notre civilisation, mais aussi la pire. Il affirme ainsi que l'IA pourrait se révéler très dangereuse si nous ne nous préparons pas aux risques potentiels liés à son utilisation (Francis, 2019). Pour terminer ce travail et réagir à ces inquiétudes, nous souhaitons proposer une série de recommandations concernant l'utilisation de la génération automatique de textes en presse écrite, dérivées des observations faites tout au long de ce mémoire. Ces propositions s'adressent à la fois aux entreprises médiatiques, aux sociétés de génération automatique de textes et aux représentants de la déontologie journalistique :

- Pour que les générateurs automatiques de textes ne représentent pas une menace pour le journalisme humain, ils devraient être considérés comme une aide pour les journalistes, et non comme un moyen de les remplacer. Dans cette optique, les logiciels leur permettraient de dégager du temps pour réaliser du contenu créatif, de l'investigation et accomplir leur mission de quatrième pouvoir.
- Pour ne pas entraver la qualité des récits journalistiques, les textes produits automatiquement devraient toujours s'adapter au sujet traité. Ils devraient également éviter les structures très répétitives et les formulations trop similaires d'articles en articles de façon à ne pas susciter de perceptions négatives ou frustrantes auprès des lecteurs.
- Dans un souci de transparence, les articles générés automatiquement devraient toujours préciser avoir été générés par des algorithmes. Les sources des données sur lesquelles les logiciels se basent devraient également être indiquées.
- En toute situation, les logiciels de génération automatique de textes devraient appliquer les règles de la déontologie journalistique, qui s'appliquent à toute production journalistique. Cet aspect peut être mis en place par programmation ou grâce à une relecture humaine.
- Un cadre juridique clair devrait être établi rapidement concernant l'utilisation de l'IA. En outre, les entreprises médiatiques devraient se préparer à assumer les conséquences d'articles générés automatiquement dont le contenu serait incorrect, inapproprié ou offensant.

Les journalistes, eux, devront être conscients des menaces que pose l'IA et prendre avantage de leur plus-value de manière à conserver leur pilier de quatrième pouvoir. Ils devront se montrer conscients des limites des algorithmes de façon à en tirer profit et à ne pas entrer en compétition avec eux. De façon plus générale, les journalistes devront continuellement rester attentifs aux dangers potentiels que représente l'utilisation de l'IA dans notre société.

Enfin, et pour conclure ce travail, nous souhaitons élargir les angles de futures recherches portant sur l'IA en journalisme. Cette étude s'est limitée à l'utilisation des générateurs automatiques de textes dans le domaine de la presse écrite. L'automatisation, cependant, concerne également la création de vidéos, qui est actuellement en plein essor et qui a notamment été popularisée par la plateforme Wibbitz (Alcaraz, 2016). Les services de Wibbitz sont notamment utilisés par *USA Today* (Alcaraz, 2016 autres), *Le Figaro* en France (Hughes, 2018) et par certaines rédactions belges, telles qu'*ELLE Belgique* (entretien du 20 janvier 2020 avec Marie Guérin, rédactrice en chef d'*ELLE Belgique*) ou encore le groupe Rossel⁸¹. En outre, en Asie, l'IA en journalisme se développe à un tout autre niveau avec l'utilisation de robots humanoïdes. En 2018, le premier robot s'apprêtait à présenter le journal télévisé au Japon (Libotte, 2018). La Chine a rapidement suivi le pas puisque depuis fin 2018, une IA nommée Qiu Hao présente également le journal télévisé (Marchand, 2018). Si ces exemples de robots humanoïdes vont au-delà de notre cadre d'étude, ils démontrent que les logiciels d'IA dans le domaine du journalisme ne se cantonnent déjà plus à la génération automatique de textes. Dès lors, dans de futures recherches, il serait pertinent d'étudier les enjeux et les questions juridiques et éthiques qui découlent de ces technologies particulières. Il serait également intéressant de se pencher sur leurs conséquences pour l'avenir du journalisme humain. L'existence de ces nouveaux logiciels renforce, par ailleurs, la nécessité que les propositions élaborées ci-dessus soient prises en compte et analysées attentivement par les autorités compétentes. Enfin, la portée de ces nouvelles réalités justifierait également que notre travail soit poursuivi et élargi à un contexte plus large que celui de la presse écrite.

⁸¹ Sources : <https://www.offremedia.com/le-groupe-rossel-choisit-la-plateforme-wibbitz-pour-faciliter-la-production-de-ses-contenus-videos>
<https://www.cbnews.fr/conseil/image-rossel-choisit-wibbitz-29880>

BIBLIOGRAPHIE

Ouvrages et chapitres d'ouvrages

- Asimov, I. (1983). *The complete robot*. London: Grafton.
- Bastin, G. (2018). *Petit lexique du Journalisme*. Fontaine: Presses Universitaires de Grenoble, PUG.
- Boden, M. (2004). *The Creative Mind: Myths and Mechanisms*. London: Routledge.
- Bourdieu, P. (1996). *Sur la télévision*. Genève: Liber.
- Collmann, J., & Matei, S. (2016). *Ethical reasoning in big data*. Washington: Springer.
- Danlos, L. (1987). *The linguistic basis of text generation*. Cambridge: Cambridge University Press.
- Deng, L., & Liu Y. (Eds.). (2018) *Deep Learning in Natural Language Processing*. Singapore: Springer.
- Einav, G. (2015). "Back to the future": Consumer and Business Trends in the digital age. In Einav, G. (Ed.), *The New World of Transitioned Media* (pp. 65–80). Springer. <https://doi.org/10.1007/978-3-319-09009-2>
- Frayn, M. (1965). *The tin men*. London: William Collins Sons & Company Ltd.
- Graefe, A. (2016). *Guide to automated journalism*. Columbia: Columbia Journalism School.
- Latar, N. (2015). The robot journalist in the age of social physics. In Einav, G. (Ed.), *The New World of Transitioned Media* (pp. 65–80). Springer. <https://doi.org/10.1007/978-3-319-09009-2>
- Latar, N. (2018). *Robot Journalism. Can Human Journalism Survive?* New Jersey: World Scientific.
- Maigret, É. (2015). Chapitre 7 - Mc Luhan et le déterminisme technologique. In Maigret, É., *Sociologie de la communication et des médias* (pp.95-106). Paris: Armand Colin.
- McNair, B. (2005). What is journalism? In de Burgh, H., *Making Journalists. Diverse Models, Global Issues* (pp.25-43). New York: Routledge.

- Neveu, É. (2009) *Sociologie du journalisme*. Paris: La Découverte.
- Pereira, F. (2007). *Creativity and Artificial Intelligence. A Conceptual Blending Approach*. Berlin: De Gruyter Mouton.
- Reiter, E., & Dale, R. (2000). *Building Natural Language Generation Systems*. Cambridge: Cambridge University Press.
- Ringoot, R. (2014). *Analyser les discours de presse*. Paris: Armant Colin.
- Ruskin A. (2018). Big data and advanced analytics. In Latar, N. (2018). *Robot Journalism. Can Human Journalism Survive?* New Jersey: World Scientific.
- Shanahan, M. (2015). *The technological singularity*. Cambridge, Massachusetts: The MIT Press.
- Thurman, N. (2019). Computational Journalism. In Hanitzsch (Eds.) (2019) *The Handbook of Journalism Studies, Second Edition*. New York: Routledge.
- Winograd, T. (1991). Thinking machines: Can there be? Are we? In Sheehan, J.J., & Sosna, M. (Eds.), *The Boundaries of Humanity: Humans, Animals, Machines* (pp. 199-224). Berkeley: University California.

Articles scientifiques

- Ali, W., & Hassoun, M. (2019). Artificial Intelligence and Automated Journalism: Contemporary Challenges and New Opportunities. *International Journal of Media, Journalism and Mass Communications*, 5(1), pp.40-49.
<http://dx.doi.org/10.20431/2454-9479.0501004>
- Andrieu, B. (2011). Quelle alternative à la domination rhétorique de la technophobie ? *Australian Journal of French Studies*, 48(2), pp.215-226. Retrieved from <https://search.proquest.com/docview/881739668/fulltextPDF/D74E00743A4B4E4A/PQ/1?accountid=14630>
- Beliën, J., Goossens, D., & Van Reeth, D. (2017). Optimization modeling for analyzing fantasy sports games. *Information Systems and Operational Research*, 5(4), pp.275-294. <https://doi.org/10.1080/03155986.2017.1279899>
- Benbouzid, B., & Cardon D. (2018). Machines à prédire. *Réseaux*, 211(5), pp.11-33.
<https://doi.org/10.3917/res.211.0009>

- Benson, R. (2006). News Media as a “Journalistic Field”: What Bourdieu Adds to New Institutionalism, and Vice Versa. *Political Communication*, 23(2), pp.187-202. <https://doi.org/10.1080/10584600600629802>
- Bertrand, C. (1993). Les M*A*R*S* en Europe ou les moyens d’assurer la responsabilité sociale. *Communication et langages*, 97, pp.22-30. <https://doi.org/10.3406/colan.1993.2450>
- Boden, M. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, 103(1-2), pp. 347-356. [https://doi.org/10.1016/S0004-3702\(98\)00055-1](https://doi.org/10.1016/S0004-3702(98)00055-1)
- Boyadjian, J., Olivesi, A., & Velcin, J. (2017). Le web politique au prisme de la science des données : Des croisements disciplinaires aux renouvellements épistémologiques. *Réseaux*, 204(4), pp.11-27. <https://doi.org/10.3917/res.204.0009>
- Bruns, A. (2019). Filter Bubble. *Internet Policy Review*, 8(4), pp.1-14. <https://doi.org/10.14763/2019.4.1426>
- Carlson, M. (2015). The Robotic Reporter. *Digital Journalism*, 3, pp.416-431. <https://doi.org/10.1080/21670811.2014.976412>
- Clerwall, C. (2014). Enter the Robot Journalist. *Journalism Practice*, 8(5), pp.519-531. <https://doi.org/10.1080/17512786.2014.883116>
- Coddington, M. (2015). Clarifying Journalism’s Quantitative turn. *Digital Journalism*, 3(3), pp.331-348. <https://doi.org/10.1080/21670811.2014.976400>
- Cohen, S., Hamilton, J. T., & Turner, F. (2011). Computational journalism. *Communications of the ACM*, 54(10), pp.66–71. <https://doi.org/10.1145/2001269.2001288>
- Cohen, N. (2015). From Pink Slips to Pink Slime: Transforming Media Labor in a Digital Age. *The Communication Review*, 18(2), pp.98-122. <https://doi.org/10.1080/10714421.2015.1031996>
- Cohen, P. (2016). Harold Cohen and AARON. *AI Magazine*, 37(4), pp.63-66. Retrieved from <https://go.gale.com/ps/i.do?p=ITOF&u=belgrca&id=GALE%7CA477085928&v=2.1&it=r>
- Cox, M. (2000). The Development of Computer Assisted Reporting. *Paper presented at the Newspaper Division, Association for Education in Journalism and Mass Communication, Southeast Colloquium, University of North Carolina, Chapel Hill*, pp.1-22. Retrieved from

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.631.6220&rep=rep1&type=pdf>

- Damper, R. (2006). The logic of Searle's Chinese room argument. *Mind March*, 16, pp.163-183. <https://doi.org/10.1007/s11023-006-9031-5>
- Danlos, L. (2010). Écriture automatique, *Les Cahiers de l'INRIA - La Recherche, INRIA*, 443, pp.1-3. Retrieved from <https://hal.inria.fr/inria-00511267/document>
- Diakopoulos, N., & Koliska M. (2017). Algorithmic Transparency in the News Media. *Digital Journalism*, 5(7), pp. 809-828. <https://doi.org/10.1080/21670811.2016.1208053>
- Dörr, K. (2016). Mapping the field of algorithmic journalism. *Digital Journalism*, 4(6), pp.700-722. <https://doi.org/10.1080/21670811.2015.1096748>
- Dörr, K., & Hollnbuchner, K. (2017). Ethical Challenges of Algorithmic Journalism. *Digital Journalism*, 6(4), pp. 404-419. <https://doi.org/10.1080/21670811.2016.1167612>
- Flew, T., Spurgeon, C., Daniel, A., & Swift, A. (2012). The promise of computational journalism. *Journalism Practice*, 6(2), pp.157–171. <https://doi.org/10.1080/17512786.2011.616655>
- Franklin, B. (2008). The Future of Newspapers. *Journalism Practice*, 2(3), pp.306-317. <https://doi.org/10.1080/17512780802280984>
- Gioti, A. (2020). From Artificial to Extended Intelligence in Music Composition. *Organized sound*, 25(1), pp.25-32. <https://doi.org/10.1017/S1355771819000438>
- Gonzales, P. (1996). Production journalistique et contrat de lecture : autour d'un entretien avec Eliseo Verón. *Quaderni*, 29, pp. 51-59. <https://doi.org/10.3406/quad.1996.1157>
- Goutefangea, P. (2016). Isaac Asimov : les (quatre) « lois de la robotique » et l'échange de parole. *hal-01397948f*. Retrieved from <https://hal.archives-ouvertes.fr/hal-01397948/document>
- Graefe, A., Haim, M., Haarmann, B., & Brosius, H. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, 19(5), pp.595-610. <https://doi.org/10.1177/1464884916641269>
- Gürses, F. (2015). Democracy, Citizenship and Local Press: The Case of the USA "Local Press in Boulder". *Advances in Journalism and Communication*, 3, pp.19-32. <http://dx.doi.org/10.4236/ajc.2015.31003>

- Hayes, A., Singer, J., & Ceppos, J. (2007). Shifting Roles, Enduring Values: The Credible Journalist in a Digital Age. *Journal of mass media ethics*, 22(4), pp.262-279. <https://doi.org/10.1080/08900520701583545>
- Kim, D., & Kim, S. (2017). Newspaper companies' determinants in adopting robot journalism. *Technological Forecasting & Social Change*, 117, pp.184-195. <http://dx.doi.org/10.1016/j.techfore.2016.12.002>
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343, pp.1203-1205. <https://doi.org/10.1126/science.1248506>
- Lewis, S., Sanders, A., & Carmody, C. (2019). Libel by Algorithm? Automated Journalism and the Threat of Legal Liability. *Journalism & Mass Communication Quarterly*, 96(1), pp.60-81. <https://doi.org/10.1177/1077699018755983>
- Matsumoto, R., Hideki, N., Tatsuya H., & Yasuo K. (2007). Journalist robot: Robot system making news articles from real world. *Proceedings of the 2007 IEEE/RSJ, international conference on intelligent robots and systems*, San Diego.
- Mercier, A., & Pignard-Cheynel, N. (2014). Mutations du journalisme à l'ère du numérique : un état des travaux. *Revue française des sciences de l'information et de la communication*, 5, pp.1-43. <https://doi.org/10.4000/rfsic.1097>
- Milosavljević, M., & Vobič, I. (2019). Human Still in the Loop. *Digital Journalism*, 7(8), pp. 1098-1116. <https://doi.org/10.1080/21670811.2019.1601576>
- Miroshnichenko, A. (2018). AI to Bypass Creativity. Will Robots Replace Journalists? (The Answer Is "Yes"). *Information*, 9(7), pp.183-203. <https://doi.org/10.3390/info9070183>
- Montal, T., & Reich, Z. (2017). I, robot. You, journalist. Who is the author? *Digital Journalism*, 5(7), pp.829-849. <https://doi.org/10.1080/21670811.2016.1209083>
- Neveu, E. (2010). News without journalists. Real threat or horror story? *Brazilian Journalism Research*, 6(1), pp.29-54. <https://doi.org/10.25200/BJR.v6n1.2010.225>
- Ombelet, P., Kuczeraway, A., & Valcke, P. (2015). Supervising automated journalists in the newsroom: liability for algorithmically produced stories. *Revue du droit des technologies de l'information*, 61, pp.5-19. Retrieved from https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1711969&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US&fromSitemap=1

- Örnebring, H. (2010). Technology and journalism-as-labour: Historical perspectives. *Journalism*, 11(1), pp.57-74. <https://doi.org/10.1177/1464884909350644>
- Plaisance, P. (2007). Transparency: An Assessment of the Kantian Roots of a Key Element in Media Ethics Practice. *Journal of Mass Media Ethics*, 22(2-3), pp.187-207. <https://doi.org/10.1080/08900520701315855>
- Ponton, C. (1997). Génération automatique de textes : 30 ans de réalisations. *Génération Automatique de textes GAT'97*, pp.1-14. Retrieved from https://www.researchgate.net/publication/253343211_GENERATION_AUTOMATIQUE_DE_TEXTES_30_ANS_DE_REALISATIONS
- Searle, J. (1990). Is the brain a mind computer program? *Scientific American*, 3(3); pp.417-458. Retrieved from <https://www.jstor.org/stable/24996641>
- Shannon, C. (1948). A Mathematical Theory of Communication. *Mobile Computing and Communication Review*, 5(1), pp.3-55. Retrieved from <http://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>
- Smyrnaiois, N. (2016). L'effet GAFAM, stratégies et logiques de l'oligopole de l'internet. *Communication & langages*, 188, pp.61-83. <https://doi.org/10.4074/S0336150016012047>
- Thurman, N., Doerr K., & Kunert, J. (2017). When reporters get hands-on with robot-writing. *Digital Journalism*, 5(10), pp.1240-1259. <http://dx.doi.org/10.1080/21670811.2017.1289819>
- Van Dalen, A. (2012). The algorithms behind the headlines. *Journalism Practice*, 6(5-6), pp.648-658. <https://doi.org/10.1080/17512786.2012.667268>
- Wu, Y. (2019). Is Automated Journalistic Writing Less Biased? An Experimental Test of Auto-Written and Human-Written News Stories. *Journalism Practice*, DOI: 10.1080/17512786.2019.1682940
- Wu, S., Tandoc, E., & Salmon, C. (2019). When Journalism and Automation Intersect: Assessing the Influence of the Technological Field on Contemporary Newsrooms. *Journalism Practice*, 13(10), pp.1238-1254. <https://doi.org/10.1080/17512786.2019.1585198>
- Zarouali, B., Van den Broeck, E., Walrave, M., & Poels, K. (2018). Predicting consumer responses to a chatbot on Facebook. *Cyberpsychology, behavior, and social networking*, 21(8) p.1-7. <https://doi.org/10.1089/cyber.2017.0518>

Articles de presse

- Alcaraz, M. (2016). Wibbitz, la start-up qui transforme des mots en vidéos. *Les Échos*. Retrieved 30 March 2020 from <https://www.lesechos.fr/2016/08/wibbitz-la-start-up-qui-transforme-des-mots-en-vidéos-234751>
- Andreasen, N. C. (2014). Secrets of the creative brain. *The Atlantic*. Retrieved 22 March 2020 from <https://www.theatlantic.com/magazine/archive/2014/07/secrets-of-the-creative-brain/372299/>
- Azar, J. (2014). Quakebot, le robot-journaliste qui rédige plus vite que son ombre. *RTBF Info*. Retrieved 25 March 2020 from https://www.rtb.be/info/dossier/chroniques/detail_quakebot-le-robot-journaliste-qui-redige-plus-vite-que-son-ombre-joyce-azar?id=8232119
- Bronner, L. (2015). Des robots au « Monde » pendant les élections départementales ? Oui... et non. *Le Monde*. Retrieved 28 March 2020 from https://www.lemonde.fr/le-monde/article/2015/03/23/des-robots-au-monde-pendant-les-elections-departementales-oui-et-non_5995670_4586753.html
- Can, Y. (2015). China's Tencent Develops a Robot Journalist to Write News Stories. *People's Daily Online*. Retrieved 25 March 2020 from <http://en.people.cn/n/2015/0911/c90000-8949019.html>
- Falcinelli, S. (2019). RTBF et « putaclic » : jusqu'où irons-nous pour vous faire cliquer ? *RTBF Info*. Retrieved 25 March 2020 from https://www.rtb.be/info/article/detail_rtb-et-putaclic-jusqu-ou-irons-nous-pour-vous-faire-cliquer?id=10148957
- Finley, K. (2015). This news-writing bot is now free for everyone. *Wired*. Retrieved 6 April 2020 from <https://www.wired.com/2015/10/this-news-writing-bot-is-now-free-for-everyone/>
- Francis, D. (2019). Artificial Intelligence. The next nuke? *The American Interest*. Retrieved 12 April 2020 from <https://www.the-american-interest.com/2019/05/24/artificial-intelligence-the-next-nuke/>
- Gautier, C. (2018). Un an avec un robot. *Alter Echos*. Retrieved 10 April 2020 from <https://www.alterechos.be/un-an-avec-un-robot/>
- Glad, V. (2015). Ce journaliste politique du Monde qui écrit aussi pour la SNCF. *Libération*. Retrieved 23 March 2020 from <http://an-2000.blogs.liberation.fr/2015/03/23/robot-journaliste-le-monde/>

- Godwin, M. (2007). Superhuman imagination. *Reason*. Retrieved 25 March 2020 from <https://reason.com/2007/05/04/superhuman-imagination/>
- Goldberg, S. (2013). Robot writers and the digital age. *American Journalism Review*. Retrieved 25 March 2020 from <https://ajr.org/2013/11/25/computer-might-replace-robot-journalism-digital-age/>
- Gorzelany, J. (2015). Volvo Will Accept Liability for Its Self-Driving cars. *Forbes*. Retrieved 10 April 2020 from <https://www.forbes.com/sites/jimgorzelany/2015/10/09/volvo-will-accept-liability-for-its-self-driving-cars/#37a126ef72c5>
- Hof, R.D. (2013). Deep Learning. *Mit Technology Review*. Retrieved 26 March 2020 from <https://www.technologyreview.com/s/513696/deep-learning/>
- Keohane, J. (2017). What news-writing bots mean for the future of journalism. *Wired*. Retrieved 23 March 2020 from <https://www.wired.com/2017/02/robots-wrote-this-story/>
- Knight, W. (2016). AI winter isn't coming. *Technology Review*. Retrieved 21 March 2020 from <https://www.technologyreview.com/s/603062/ai-winter-isnt-coming/>
- Lambert, X. (2017). 500 000 clics sur des articles générés par des robots : les journalistes doivent-ils craindre l'intelligence artificielle ? *RTBF Info*. Retrieved 25 March 2020 from https://www.rtbf.be/info/medias/detail_850-articles-du-washington-post-crees-par-des-robots-les-journalistes-doivent-ils-craindre-l-intelligence-artificielle?id=9711388
- Levy, S. (2012). Can an Algorithm Write a Better News Story Than a Human Reporter ? *Wired*. Retrieved 6 April 2020 from <https://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/>
- Levy, S. (2013). How Ray Kurzweil Will Help Google Make the Ultimate AI Brain. *Wired*. Retrieved 21 March 2020, from <https://www.wired.com/2013/04/kurzweil-google-ai/>
- Libotte, A. (2018). Erica devrait être le premier robot à présenter un journal télévisé au Japon. *RTBF Info*. Retrieved 2 April 2020 from https://www.rtbf.be/info/medias/detail_et-si-le-journal-televisé-etait-presenté-par-un-robot?id=9837929

- Marchand, L. (2018). En Chine, le JT est présenté par une Intelligence Artificielle. *Les Échos*. Retrieved 2 April 2020 from <https://www.lesechos.fr/tech-medias/medias/en-chine-le-jt-est-presente-par-une-intelligence-artificielle-146635>
- Marr, B. (2015). Can Big Data Algorithms Tell Better Stories than Humans ? *Forbes*. Retrieved 25 March 2020 from <https://www.forbes.com/sites/bernardmarr/2015/07/22/can-big-data-algorithms-tell-better-stories-than-humans/#4cf1cdbf14b5>
- Mc Farland, M. (2015). Associated Press looks to expand its automated stories program following successful launch. *The Washington Post*. Retrieved 6 April 2020 from <https://www.washingtonpost.com/news/innovations/wp/2015/01/29/associated-press-looks-to-expand-its-automated-stories-program-following-successful-launch/>
- Metz, R. (2017). Why Google's AI can write beautiful song but still can't tell a joke. *MIT Technology Review*. Retrieved 22 March 2020 from <https://www.technologyreview.com/s/608777/why-googles-ai-can-write-beautiful-songs-but-still-cant-tell-a-joke/>
- Moreira, E. (2017). Quand l'intelligence artificielle invente un langage incompréhensible par l'homme. *Les Échos*. Retrieved 21 March 2020 from <https://www.lesechos.fr/2017/06/quand-lintelligence-artificielle-invente-un-langage-incomprehensible-par-lhomme-174329>.
- Podolny, S. (2015). If an algorithm wrote this, how would you even know? *The New York Times*. Retrieved 25 March 2020 from <https://www.nytimes.com/2015/03/08/opinion/sunday/if-an-algorithm-wrote-this-how-would-you-even-know.html>
- Rauline, P. (2019). « La crise de la presse locale a un coût », estime une chercheuse de la Brookings Institution. *Les Échos*. Retrieved 19 April 2020 from <https://www.lesechos.fr/tech-medias/medias/la-crise-de-la-presse-locale-a-un-cout-economique-et-politique-estime-une-chercheuse-de-la-brookings-institution-1150862>
- Somers, J. (2017). Is AI riding a one trick pony? *MIT Technology Review*. Retrieved 21 March 2020 from <https://www.technologyreview.com/s/608911/is-ai-riding-a-one-trick-pony>
- The New York Times (2015). Did a Human or a Robot Write This? *The New York Times*. Retrieved 8 April 2020 from https://www.nytimes.com/interactive/2015/03/08/opinion/sunday/algorithm-human-quiz.html?smid=pl-share&_r=0

Warsztacki, S. (2017). Un robot à la rédac. *Alter Echos*. Retrieved 10 April 2020 from <https://www.alterechos.be/un-robot-a-la-redac/>

WashPostPR (2017). The Washington Post leverages automated storytelling to cover high school football. *The Washington Post*. Retrieved 6 April 2020 from <https://www.washingtonpost.com/pr/wp/2017/09/01/the-washington-post-leverages-heliograf-to-cover-high-school-football/>

Sources universitaires

Loupe, G. (2018). *INFO8006-1 : Introduction to Artificial Intelligence*. Cours ex-cathedra, Université de Liège, Liège.

Stone, M. (2014). *Big Data for Media*. Report of the Reuters Institute for the Study of Journalism. University of Oxford, Oxford.

Taillefer, H. (2009). *L'intelligence artificielle comme figure de la dystopie dans « Nineteen eighty four », de George Orwell, « Le dépeupleur », de Samuel Beckett et « Neuromancer » de William Gibson* (Master's thesis). Université du Québec à Montréal, Québec.

Targnion, P. (2019). *Les algorithmes des sites d'infodivertissement destinés aux jeunes* (Master's thesis). Université de Liège, Liège.

Sources audiovisuelles

Balbastre, G., & Kergoat, Y. (Réalisateurs). (2011). *Les nouveaux chiens de garde* [Film]. JEM Productions.

De Loupy, C. (2018). *Quand les robots deviennent écrivains* [Video File]. Retrieved 11 April 2020 from <https://www.youtube.com/watch?v=Btj7ONUzGXg>

Kahneman, D. (2017). *Remarks* [Video File]. NBER Conference Toronto: Economics of AI Conference. Toronto. Retrieved 21 March 2020 from <https://digitopoly.org/2017/09/22/Kahneman-on-ai-versus-humans/>

Searle, J. (2013). *Our shared condition consciousness* [Video File]. Retrieved 22 March 2020 from https://www.ted.com/talks/john_searle_our_shared_condition_consciousness/transcript?language=fr#t-6386

Sources Internet

- Bradshaw, P. (2016). FAQ: Do we need new ethics for computational journalism? *Online Journalism Blog*. Retrieved 15 April 2020 from <https://onlinejournalismblog.com/2014/08/16/faq-do-you-need-new-ethics-for-computational-journalism/>
- Burić, M. (n.d.). Intelligence artificielle et droits de l'homme. *Conseil de l'Europe*. Retrieved 12 April 2020 from <https://www.coe.int/fr/web/artificial-intelligence/secretary-general-marija-pejcinovic-buric>
- Bremme, L. (n.d.). Définition : Qu'est-ce que le Big Data ? *lebigdata.fr*. Retrieved 26 March 2020 from <https://www.lebigdata.fr/definition-big-data>
- Carter, J. (2013). Could robots be the writers of the future? *Techradar.com*. Retrieved 25 March 2020 from <https://www.techradar.com/news/computing/could-robots-be-the-writers-of-the-future-1141399>
- DeGeurin, M. (2018). A Startup Media Site Says AI Can Take Bias Out of News. *Vice*. Retrieved 11 April 2020 from https://www.vice.com/en_us/article/zmgza5/knowhere-ai-news-site-profile
- Dickson, B. (2017). What is Narrow, General and Super AI. *Tech Talks*. Retrieved 21 March 2020 from <https://bdtechtalks.com/2017/05/12/what-is-narrow-general-and-super-artificial-intelligence/>
- Dierickx, L. (2015) Petite histoire de la génération automatique de textes. *Ohmybox*. Retrieved 20 April from <https://www.ohmybox.info/petite-histoire-de-la-generation-automatique-de-textes/>
- Dierickx, L. (2019). Production automatisée d'informations : une ligne du temps. *Ohmybox*. Retrieved 20 April 2020 from <https://www.ohmybox.info/generation-automatique-de-textes-et-journalisme-une-ligne-du-temps/>
- Doctor, K. (2013). The newsonomics of “Little Data,” data scientists and conversion specialists. *NiemanLab*. Retrieved 21 March 2020 from <https://www.niemanlab.org/2013/10/the-newsonomics-of-little-data-data-scientists-and-conversion-specialists/>
- Ha, A. (2018). News startup Knowhere aims to break through partisan echo chambers. *TechCrunch*. Retrieved 11 April 2020 from

<https://techcrunch.com/2018/04/04/news-startup-knowhere-aims-to-break-through-partisan-echo-chambers/>

Hughes, S. (2018). Comment le Figaro a adopté la vidéo sur ses plateformes numériques. *Wibbitz*. Retrieved 29 April 2020 from <https://www.wibbitz.com/fr/blog/le-figaro-strategie-video-plateformes-numeriques/>

Kleijssen, J. (n.d.). Comment encadrer le développement de l'intelligence artificielle ? *Conseil de l'Europe*. Retrieved 19 April 2020 from <https://www.coe.int/fr/web/artificial-intelligence/directeur-jan-kleijssen>

Marshall, S. (2013). Robot reporters: A look at the computers writing the news. *Journalism.co.uk*. Retrieved 25 March 2020 from <https://www.journalism.co.uk/news/robot-reporters-how-computers-are-writing-la-times-articles/s2/a552359/>

Miller, R. (2015). AP's robot journalists' are writing their own stories now. *The Verge*. Retrieved 10 April 2020 from <https://www.theverge.com/2015/1/29/7939067/ap-journalism-automation-robots-financial-reporting>

Mullin, B. (2016). The Washington Post will use automation to help cover the election. *Poynter*. Retrieved 6 April 2020 from <https://www.poynter.org/tech-tools/2016/the-washington-post-will-use-automation-to-help-cover-the-election/>

Raynaud, P., & Didier, I. (2018). Production automatique de textes : l'IA au service des journalistes. *Institut National de l'audiovisuel*. Retrieved 11 April 2020 from <https://larevuedesmedias.ina.fr/production-automatique-de-textes-lia-au-service-des-journalistes>

Rainie, L., & Anderson, J. (2017). The future of jobs and job training. *Pew Research Center*. Retrieved 22 March 2020 from <https://www.pewresearch.org/internet/2017/05/03/the-future-of-jobs-and-jobs-training/>

TechTarget (2015). Definition, data silo. *TechTarget.com*. Retrieved 26 March 2020 from <https://searchdatamanagement.techtarget.com/definition/data-silo>

Smith, S. (2015). An NPR Reporter Raced A Machine to Write a News Story. Who won? *NPR*. Retrieved 8 April 2020 from <https://www.npr.org/sections/money/2015/05/20/406484294/an-npr-reporter-raced-a-machine-to-write-a-news-story-who-won?t=1586332887739>

Weinberger, D. (2009) Transparency is the New Objectivity. *Joho the Blog*. Retrieved 20 April 2020 from <https://www.hyperorg.com/blogger/2009/07/19/transparency-is-the-new-objectivity/>

Ce travail a été rédigé selon les normes bibliographiques de l'American Psychological Association (APA).

ANNEXES

Annexe 1 : Interview de Marie Guérin **Rédactrice en chef d'ELLE BELGIQUE**

Entretien réalisé le 20 janvier 2020

- **Retranscription⁸²** -

Selon vous, intelligence artificielle et journalisme vont-ils de pair ?

Oui, je pense que l'intelligence artificielle, en journalisme comme ailleurs, peut être un outil incroyable pour se développer, pour être plus précis, plus ciblé, pour décupler ses capacités. Je pense que ça fonctionne tout à fait, bien sûr.

Vous avez déjà fait appel à Textbroker.fr, une plateforme où des rédacteurs rédigent pour vous des articles sur un sujet donné. Pourquoi ?

On a déjà utilisé cette plateforme parce que notre grand projet en 2019, et aussi en 2020, c'était de générer des articles plutôt généralistes sur des grandes thématiques, du genre « comment choisir son mascara ? ». On voulait avoir des articles très larges qui rassemblaient un maximum de mots-clés et qui pouvaient être référencés en position zéro sur Google⁸³. Ce sont des articles qui sont très bien référencés. Mais seulement, on n'avait pas assez de journalistes pour le faire et donc on s'est dit que pour faire cette base d'articles de socle, on allait utiliser les services de Textbroker.fr. C'est une plateforme où des gens écrivent pour vous. On leur donne 150 mots-clés qui doivent être dedans. C'est comme des machines, ils vont les inclure dans leur article. Après, on doit quand même repasser dedans, les corriger, mais on sent quand même la patte humaine. Parce que la différence avec l'intelligence artificielle je trouve, c'est que dans les magazines féminins, on a vraiment un ton qui est très proche de la lectrice, on essaie de mettre de l'humour et des choses comme ça, ce qu'on a moins dans la presse généraliste, où c'est très factuel. Et donc, ça, c'est quelque chose que l'on n'a pas avec Textbroker.fr. Donc on doit quand même repasser après. Mais par contre, cette capacité à faire une phrase en mettant tous les mots-clés, ça c'est génial. On l'utilise plutôt pour réaliser des articles rébarbatifs où on doit compiler. Pour faire un article sur comment mettre du mascara, il n'y a pas besoin d'aller faire un interview. On peut aller voir

⁸² Toutes les retranscriptions de ce mémoire sont des retranscriptions *verbatim*, réalisées mot à mot.

⁸³ Premier lien proposé par le moteur de recherche Google lors d'une requête.

quinze autres articles là-dessus, on va sélectionner, donc ça, ce n'est pas du journalisme. Ce n'est pas super excitant.

Vous programmez de continuer à l'utiliser ?

Oui, je pense. Parce que de toute façon la technique va s'affiner et un jour on arrivera à des articles bien léchés. Mais je pense qu'on est la seule rédaction féminine à réagir comme ça.

Les générateurs automatiques de textes peuvent-ils être considérés comme une menace pour le journalisme ?

Je sais que la profession considère ça comme une menace. À titre personnel, j'estime que si quelqu'un se sent menacé, c'est qu'il n'a pas foi en ses compétences. Moi, je crois vraiment en la technologie et j'estime que ce sera toujours une aide. Et si quelqu'un croit qu'il sera remplacé à 100 %, c'est que ce qu'il fait n'est pas suffisamment optimal. Moi, je suis persuadée qu'on aura toujours une plus-value et qu'on travaillera de concert. La technologie va juste nous apporter un confort. Par exemple, tous ces articles embêtants qu'on n'a pas envie de faire, ils les feront pour nous. Et on va pouvoir aller faire des portraits et des interviews créatives que la machine ne pourra pas faire. Donc moi, je vois plutôt ça comme un allié. De toute façon ça va arriver. Les journalistes pourront se battre et dire tout ce qu'ils veulent, on arrivera à ça.

C'est donc plus une opportunité qu'une menace ?

Oui, tout à fait. Parce que si la lectrice veut savoir exactement comment est fabriqué un mascara, elle n'a pas besoin de moi, c'est une information factuelle. Mais par contre, si elle veut lire les propos d'une femme dont c'est le travail, qui raconte avec passion ce que c'est et avec une vraie direction d'interview, avec un très chouette portrait émouvant et touchant, ça je peux mieux le faire. Mais je pense que c'est moins facile pour les dépêches d'information, ça pourrait se systématiser et il n'y aurait plus de journalistes. Mais les grandes enquêtes de fond, il y a un vrai travail journalistique. On aura les infos généralistes écrites par des robots et puis on aura des choses plus créatives faites par les journalistes.

Quelle peut être la plus-value d'un journaliste par rapport à un générateur automatique de textes ?

On va faire des choses différentes. L'intelligence artificielle va être factuelle. Le journaliste, il va apporter sa plume, sa sensibilité. On dit toujours que le journalisme doit être objectif.

Moi, je n'y ai jamais cru parce que nous sommes des êtres humains et qu'on met toujours notre sensibilité dans les articles que nous écrivons. Même si on essaie d'être neutre, on va quand même y mettre nos touches, nos références ou notre humour. Même si un jour les intelligences artificielles sont capables de faire de l'humour, etc., nous, on aura toujours la créativité. On fera toujours mieux qu'une intelligence artificielle ou avant elle. C'est ça la différence. Et aussi, les gens voudront avoir l'avis de Juliette Debruxelles sur tel événement ou voudront avoir sa touche, sa façon d'interroger. Ce sera toujours différent. Donc on sera plus dans des papiers engagés, ou plus sur l'émotion, et les articles plus factuels seront réservés aux intelligences artificielles.

Annexe 2 : Interview d'Émilie Mascia **Rédactrice en chef d'ELLE à table Belgique**

Entretien réalisé le 20 janvier 2020

- Retranscription -

Intelligence artificielle et journalisme vont-ils de pair ?

De base, pour moi, ce sont deux mots sont assez opposés. Parce que le journalisme c'est une critique, c'est une vision du monde par une personne. Pour moi, le journalisme n'est jamais vraiment objectif puisqu'il y a toujours quelqu'un derrière la plume. Donc c'est une analyse, c'est quelqu'un qui va amener son ressenti dans un texte, qui va trouver des idées, qui va être créatif. Donc on pourrait dire que dans un sens, l'intelligence artificielle, ça va très loin maintenant. Google est une intelligence artificielle. Et donc, nous, on doit réagir, on doit travailler nos articles pour répondre à Google et pour être mieux référencés en SEO. Donc on a notre côté créatif, mais on est un peu bridés par le fait qu'il faut de plus en plus répondre à des règles qui sont fournies par des robots. Pour le web, maintenant, on ne peut plus ne pas faire attention à ça. Donc moi, je me suis créé une petite *to-do list*. Il faut toujours un titre, une introduction, un texte de minimum 300 mots, sinon on n'est pas vraiment référencé. Il faut essayer d'avoir des mots-clés à répétition dans le texte. Il faut bien renommer le texte alternatif des photos. Si on décrit bien la photo, on sera mieux référencé dans Google. Il y a d'innombrables étapes à remplir. Donc aujourd'hui, faire un article web, ce n'est plus seulement copier-coller un texte qu'on avait fait pour le magazine papier et le mettre en ligne. Non, il faut redimensionner les photos, il faut penser à tout ce qui est SEO, il faut que le nombre de mots soit parfait, il faut retravailler les titres pour que ce soit accrocheur et en même temps bien référencé. Il y a tout un travail qui est beaucoup plus poussé qu'avant. C'est presque un nouveau métier. En tout cas, on voit que les journalistes qui sont là depuis 15 ans ont un peu plus de mal à s'adapter à cette nouvelle façon de travailler.

Que pensez-vous des générateurs automatiques de textes ?

Il y aura plusieurs positionnements : le positionnement vis-à-vis de Google, vis-à-vis du lecteur et aussi l'aspect commercial. Quand je fais un article pour le web, je me dis que je vais déjà aller chercher sur Google ce qu'il existe par rapport à cet article. En fait, on va travailler à l'envers. On a un analyste SEO qui va nous fournir une liste de mots-clés, des questions fréquentes par rapport au sujet. Donc ça, ça pourrait à la limite être fait par un robot si on lui disait qu'il y a cinquante mots-clés à insérer dans le texte, qu'il y a telles thématiques

à inclure, etc. Si l'intelligence arrive à le faire, pourquoi pas. Maintenant, quand l'analyste SEO sort une liste de mots-clés, parfois il y en a cinquante, mais il y en a qu'on ne mettrait jamais dans le texte dans la logique des choses. Et donc il faut aussi penser aux lecteurs. On avait déjà fait des commandes où on voulait des textes de 800 mots pour avoir beaucoup de données sur le site. Mais quand tu lis ça, c'est imbuvable. On ne peut pas partager ça sur Facebook, ils ne vont jamais le lire, les mêmes mots sont répétés trop de fois. Le journalisme ne va pas mourir, le robot ne va pas le remplacer dans l'écriture. Quoique... peut-être que plus tard il aura un style, une manière d'écrire, de l'humour ou des caractères qui font qu'on pourra penser que c'est un humain qui écrit. Je pense qu'on n'est pas encore à ce niveau-là. Mais je pense qu'il ne faut pas non plus être réfractaire à ces nouvelles technologies. Ça peut devenir un allié, une sorte de partenariat. Chez *ELLE*, notre but, c'est toujours d'augmenter le trafic sur le site internet dans l'objectif final de gagner de l'argent. La presse, pour l'instant, c'est difficile. Donc le web c'est un peu le dernier recours où on peut générer de l'argent. Si travailler avec l'intelligence artificielle nous permet d'arriver à ce but-là, bien sûr, nous sommes partants.

Si on vous offrait un robot qui génère automatiquement des articles, accepteriez-vous ?

Ça dépend si c'est moi qui suis payée et pas le robot. Je pense qu'il y aurait quand même un travail de remise en forme derrière. Nous, on serait plus dans une position de stratégie, de décider de quels sujets traiter etc., comme s'il n'y avait plus que des éditeurs et que les exécutants n'existaient plus. Ce sera peut-être dans un futur proche. Dans vingt ans, on verra.

Les générateurs automatiques de textes pourraient-ils faciliter le travail d'un journaliste ?

Faciliter, pas vraiment. Parce que c'est parfois très prenant de corriger un texte. Parfois, c'est plus rapide de l'écrire entièrement. Et je pense que le journaliste a toujours une plus-value à apporter, jusqu'à preuve du contraire. On peut parler de tout ce qui est émotion. Qu'est-ce qui fait qu'un article marche mieux qu'un autre ? Ça va être en fonction du vécu d'une personne, de l'émotion. Je pense qu'un robot qui sortirait des articles sur des choses qui ne sont pas forcément en lien avec ce que les gens vivent, ça passerait peut-être très vite aux oubliettes.

L'élément humain est-il important dans un article ?

Oui, moi j'en suis convaincue. Après, c'est vrai que maintenant il y a peut-être beaucoup de gens qui veulent juste l'information et qui n'ont pas vraiment d'attrait pour la façon d'écrire et qui ne regardent pas les articles d'un œil littéraire. Mais tout dépend du type de média. Si c'est juste de l'information, quand on parle de *news*, peut-être que ce serait intéressant d'avoir des robots qui lâchent les informations, un peu comme les agences de presse avec les dépêches. D'ailleurs, j'ai déjà un peu l'impression que c'est quasiment des robots qui font ces dépêches. C'est tellement mal écrit qu'on ne peut pas copier-coller ça tel quel. Et puis les journalistes humains ont des qualités que les robots n'ont pas : le style d'écriture, l'aspect critique, la réflexion, la vérification des données. Dans le domaine *lifestyle*, ces éléments-là sont très importants. Après, pour des petites actualités ou des tendances, pourquoi pas, peut-être qu'un robot pourrait bien mettre ça en forme. Mais il faudra quand même toujours des humains derrière.

Annexe 3 : Interview de Nadia Chater

Journaliste aux *Éditions L'Avenir* et coordinatrice de *Deuzio*

Entretien réalisé le 30 mars 2020

- Retranscription -

Selon vous, journalisme et intelligence artificielle vont-ils de pair ?

Je n'en ai aucune idée. Non, le journalisme, c'est quelque chose qui fait appel à une intelligence réelle et pas une intelligence artificielle, donc ça ne me semble pas cohérent d'assembler les deux ensemble.

Que pensez-vous des générateurs automatiques de textes en presse écrite ?

Je ne suis pas d'accord. Je trouve ça complètement débile. Pourquoi générer du texte comme ça... Il faut une recherche, il faut un travail d'écriture. Ça ne me semble pas du tout cohérent. Ça ne peut pas s'appliquer à du journalisme. Je ne savais pas que cette technologie existait. Ça existe vraiment ?

Vous ne souhaiteriez pas que cette technologie soit utilisée aux *Éditions L'Avenir* ?

Non. Je trouve qu'on a besoin de vrais cerveaux pour écrire, même si c'est pour des comptes rendus de sport par exemple. Il faut quand même un minimum d'activité humaine, de réflexion. Non, pour moi, ça ne se justifie pas du tout.

Selon vous, l'IA représente-t-elle une menace pour le « journalisme humain » ?

Je ne suis même pas sûre parce que je pense que ça a ses limites. Je ne suis pas sûre que les gens aient envie de lire de l'information générée automatiquement. Je crois que ce n'est même pas envisageable.

Si les générateurs automatiques de textes produisaient des articles très factuels et permettaient aux journalistes de se concentrer sur des activités plus créatives ou des articles de temps long, seriez-vous toujours contre ?

Oui. Franchement, je ne vois pas du tout ce que ça viendrait faire dans ce métier. Pour moi, ça ne s'adapte pas du tout au journalisme. Tout article doit être rédigé par un être humain. Sinon, ça n'a pas de sens. Pour moi, ça ne se justifie pas. Même si c'est pour faire gagner du temps. Je pense qu'il faut dégager des moyens pour le journalisme, mais pas de cette façon-

là. Le journalisme a besoin de plus de moyens. Ces dernières années, ils ont clairement diminué les budgets. Avec Internet, on doit écrire de plus en plus vite, on a de moins en moins de temps à la réflexion et c'est vraiment un problème. On pense « clic » et pas « réflexion ». Or, quand on voit les bonnes enquêtes qui sont menées par certains journaux, on voit bien l'intérêt de dégager du temps. Il faudrait redonner un peu de temps aux journalistes, mais surtout pas grâce à une intelligence artificielle.

Quelle est, selon vous, la plus-value des journalistes par rapport aux robots ?

L'intelligence humaine, le sens de la réflexion. Aller chercher l'information où elle se trouve, creuser une information et ne pas simplement se contenter de ce que l'on entend ou ce que l'on lit. Ça, je pense que l'intelligence artificielle ne l'aura jamais. L'intelligence humaine n'est pas reproductible par une machine.

Annexe 4 : Interview de Ricardo Gutiérrez

Secrétaire général de la Fédération européenne des journalistes (FEJ) et membre du Conseil de Déontologie Journalistique (CDJ)

Entretien réalisé le 30 mars 2020

- Retranscription -

De façon générale, que pensez-vous des générateurs automatiques de textes en presse écrite ?

Nous, comme organisation représentationnelle du journalisme, on s'est évidemment intéressé à l'intelligence artificielle, d'abord par le biais de ce qu'est le robot-journalisme. Donc il y a eu des premières rédactions, comme le journal *Le Monde* à l'occasion des municipales en France, qui se sont dotées de ces outils. Et je dirais qu'on a eu une approche assez ouverte qui était de se dire, tant que ces outils sont des outils, c'est-à-dire qu'ils permettent aux journalistes de mieux faire leur travail ou qu'ils accomplissent des tâches ardues qui seraient parfois impossibles à effectuer pour un simple journaliste, nous sommes d'accord. Je vous rappelle que *Le Monde*, en couvrant les municipales, a couvert énormément de résultats. Ce serait impossible de le faire avec des journalistes. Par contre, l'intelligence artificielle et les robots générateurs de textes nous permettent de le faire, tant que c'est encadré par des humains, évidemment. Donc on n'est pas contre cette évolution technologique-là, pour autant qu'elle soit encadrée, c'est-à-dire qu'on ne doit pas laisser les robots prendre le pouvoir. Les générateurs de textes c'est très bien tant que les logiciels sont sous contrôle humain et notamment que les balises éthiques sont celles de la profession. On ne publie pas tout et n'importe quoi en tant que journaliste. Toutes les informations qu'on reçoit, on les trie, on les vérifie. Il y en a qu'on décide de ne pas publier même si elles sont vraies, parce que, par exemple, elles pourraient mettre en danger la vie de quelqu'un. On a parfois besoin de protéger les sources parce qu'on sait que ces sources vont parfois faire l'objet de représailles, par exemple. Toutes ces normes, toutes ces balises éthiques, la machine ne les a pas forcément. Donc il faut vraiment qu'il y ait un contrôle humain sur la production. La position de la FEJ, c'est qu'on n'est pas contre cette évolution-là, elle peut aider le journaliste. Et pendant qu'un robot générateur de textes accomplit des tâches banales ou basiques, ça laisse justement le temps au journaliste de faire des tâches beaucoup plus ardues, comme le journalisme d'investigation qui demande des contacts interpersonnels. Une machine ne pourra jamais faire de l'investigation. Elle peut aider l'investigateur mais elle ne peut pas faire de l'investigation à sa place parce qu'elle demande des interactions humaines et des

relations de confiance. Si une personne vous dit quelque chose parce que vous avez gagné sa confiance, elle ne le dira jamais à une machine. Il y a des choses qui continueront d'être de la prérogative humaine. Et donc, tant mieux si la machine prend en charge des tâches basiques comme collecter les résultats de foot ou des résultats électoraux. C'est des chiffres, ça n'engage pas à grand-chose de dire que le PS, dans telle commune française, a fait 25% aux dernières élections. Est-ce qu'on a besoin d'un journaliste pour dire ça ? Non. Est-ce qu'une machine peut le faire beaucoup mieux et beaucoup plus rapidement qu'un journaliste, certainement et qu'elle le fasse. Donc notre position est assez ouverte. Mais en même temps, on a des balises éthiques. On ne peut pas faire faire n'importe quoi aux générateurs automatiques de textes. Il y a des sujets beaucoup plus sensibles où la ligne éditoriale d'un journal intervient. Quand on fait une analyse politique par exemple, quand on écrit un éditorial. J'imagine mal un éditorial écrit par un générateur de textes. Parce que là, ça suppose de la discussion interne, des positionnements par rapport à des politiques précises qui ne sont pas les mêmes d'un journal à l'autre. Je dirais que pour tout ce qui demande un positionnement éthique ou politique, on continue, plus que jamais, à avoir besoin des journalistes en chair et en os. Donc les machines, très bien si elles peuvent aider à accomplir des tâches ingrates ou impossibles à tenir pour un homme, mais elles ne doivent pas remplacer les journalistes dans des registres beaucoup plus délicats qui engagent vraiment la ligne éditoriale d'un journal ou qui demandent un jugement humain. Parfois, rien n'est tout blanc ni tout noir. C'est pour ça qu'on a des conseils de déontologie en journalisme. Par rapport à n'importe quel article, on peut avoir des doutes et il est toujours bon à ce moment-là d'avoir une réflexion de type humaine et concertée. Souvent, quand il y a des questions qui se posent, le journaliste n'est pas seul. Il en discute avec son chef de service, au rédacteur en chef, il y a une discussion interne, il consulte éventuellement les instances déontologiques. Et ça, une machine ne pourra pas le faire. Une autre balise que l'on demande, c'est que quand un texte est écrit par une machine, il faut que le lecteur le sache. Donc ça doit être connu du lecteur, on ne doit pas utiliser ces machines à l'insu du lecteur. Pour le moment, il n'y a pas de loi là-dessus parce que ce sont des éléments neufs. Le Conseil de l'Europe a très vite mis en place un comité interne. Il y a toute une série de travaux en cours sur les implications de l'intelligence artificielle dans toute la société, parce qu'il y a des implications éthiques. Il y a déjà eu des recommandations, signées par les États membres du Conseil de l'Europe en 2007, sur l'implication de l'intelligence artificielle sur les droits de l'homme, comme quoi on doit protéger les droits de l'homme contre d'éventuelles dérives de l'intelligence artificielle. C'est un outil pour les journalistes. Et un outil peut être utilisé à bon escient ou à

mauvais escient. Un outil n'est pas mauvais ou bon en soi, c'est la manière dont vous l'utilisez qui est bonne ou mauvaise. Donc nous, on n'est pas contre l'outil mais on veut que ce soit encadré. Et le Conseil de l'Europe va dans le même sens. Il estime qu'il faut des régulations de l'intelligence artificielle. On ne peut pas faire tout et n'importe quoi avec des algorithmes parce qu'ils peuvent conduire à des atrocités ou des choses abominables. Donc il faut un encadrement et cet encadrement a commencé puisqu'il y a déjà des recommandations. Mais ces outils, on ne veut pas les diaboliser non plus. On trouve qu'ils peuvent être très utiles aux journalistes. Maintenant, il faut que le lecteur qui est en train de lire le sache. Et par ailleurs, qu'il y ait un encadrement humain.

Les générateurs automatiques de textes représentent-ils une menace pour le journalisme humain ?

Non, au contraire, c'est une opportunité. Faire faire à des machines des tâches ingrates, c'est justement revaloriser le journaliste dans sa mission essentielle, qui est d'être au service du citoyen, d'être un contre-pouvoir. Défendre les positions et les intérêts des citoyens face à tout qui détient pouvoir dans notre société, ça, c'est le rôle fondamental du journaliste. Tous les journalistes ne font pas ça. Ils ne sont pas en capacité de faire ça pour le moment. Ils voudraient peut-être bien être les chiens de garde de la démocratie. Moi, je n'en connais pas beaucoup qui le sont, malheureusement. Donc si ces outils permettent d'abattre les tâches ingrates et de faire en sorte que les journalistes redeviennent des chiens de garde de la démocratie, et bien tant mieux. Donc nous, on ne voit pas du tout un danger pour le journalisme, on y voit au contraire une opportunité, une chance pour les journalistes de renouer avec leurs fondamentaux, leur mission de base, qui n'est pas celle de recopier bêtement un communiqué de presse, mais de se poser des questions. Après, ça dépend comment on l'utilise. Si une entreprise de presse décide de licencier tous ses journalistes et de confier le journal à des générateurs de textes, ça ne répond absolument pas à ce que l'on veut. Finalement, les générateurs de textes, ça pourrait être une solution si on l'applique intelligemment.

Pensez-vous qu'un jour, des articles du code de déontologie journalistique concerneront les générateurs automatiques de textes ?

Pour moi, le code de déontologie s'applique déjà aux générateurs de textes automatiques. Le code de déontologie s'applique à des productions journalistiques. On ne dit pas que ces productions doivent avoir été produites par un journaliste humain. Si un journal publiait un

article créé par un générateur de textes et qu'il y avait une faute déontologique dans cet article, rien n'empêche le conseil de déontologie, aujourd'hui, de sanctionner cet article. Donc pour moi, le problème de la déontologie, c'est un faux problème. La déontologie s'applique à toute production journalistique, qu'elle soit créée par un être humain ou une machine. Il n'y a aucune différence. Un citoyen qui trouverait qu'un article n'est pas déontologique a le droit de déposer plainte au CDJ, peu importe que l'article ait été produit par une machine. Quand on dépose plainte au CDJ, on ne sait pas toujours d'ailleurs, Parfois, ce sont des articles qui ne sont pas signés. La déontologie reste la déontologie. On ne va pas faire une déontologie spéciale pour l'intelligence artificielle. La déontologie journalistique s'applique à toute production journalistique, humaine ou non humaine, peu importe. C'est la déontologie journalistique, les machines doivent s'y plier. Donc les logiciels derrière ces générateurs de textes doivent prendre en compte la dimension éthique. Mais ces règles existent déjà, il n'y a pas besoin de les adapter. Quand la loi nous dit qu'on ne doit pas citer le nom d'un mineur impliqué dans une affaire de meurtre, c'est appliqué par qui que ce soit, un journaliste humain comme un générateur de textes. Ce sont des données qu'on peut intégrer au logiciel, faire en sorte que dès qu'un mineur intervient dans un article, on ne cite pas son identité. La déontologie s'applique à tout, il ne faut pas changer les règles parce qu'il y a les générateurs de textes, c'est les mêmes règles pour tout le monde.

Annexe 5 : Interview d'André Thiel **Directeur *Process & Development* chez *SudPresse***

Entretien réalisé le 30 mars 2020

- Retranscription -

Utilisez-vous la technologie de génération automatique de textes chez *Sudpresse* ?

Oui, nous l'utilisons avec un partenaire. Nous sommes en partenariat avec la société parisienne LabSense qui écrit des textes sur base de données de toutes sortes. Il y a pas mal d'opérateurs sur le marché maintenant, mais cette société-là était déjà partenaire d'un groupe de presse en France avec lequel on a beaucoup de contacts. On travaille pour l'instant sur trois axes principaux. Le premier axe, c'est le pôle local. On écrit automatiquement des articles sur des matches qu'on ne pourrait pas couvrir avec du journalisme humain en s'appuyant sur un partenariat qu'on a avec l'Union belge de Football. On écrit un article sur chaque match organisé par l'Union belge de Football durant tout le week-end et toute la semaine. Cela signifie que plusieurs milliers de résumés de matches par semaine sont générés automatiquement. Ils sont générés dans le cadre de cet accord qu'on a avec l'Union belge qui nous donne accès à toute sa base de données. Il faut savoir qu'aujourd'hui, les feuilles de matches sont informatisées. C'est ça qui a permis, ces dernières années, d'obtenir assez vite, après la fin de la rencontre, les indicateurs principaux d'un match. À savoir le résultat, qui a marqué, à quelle minute, le nom de l'arbitre, les gens qui ont joué, les remplaçants, etc. On est donc capable de rédiger des articles qui sont assez simples puisqu'ils se basent sur de la *data*. On enrichit alors ces articles avec un autre procédé qui relève aussi de l'intelligence artificielle. Au fond, l'intelligence artificielle, dans ce cas-là, c'est essentiellement un tableau Excel qui est donné à manger à un convertisseur de textes. Et l'expertise des boîtes comme LabSense, elle est dans la sémantique et dans le fait de réussir à écrire des articles et de varier les tournures de phrases pour ne pas toujours écrire les mêmes articles sur tous les matches. C'est la valeur ajoutée de ces boîtes-là. Donc on enrichit alors les articles avec des infographies générées automatiquement sur base des mêmes données. Les infographies présentent, par exemple, la position au classement. Ça, il n'y a personne qui le fait parce que c'est hors de portée, on n'arriverait clairement pas à couvrir plus de 6000 articles par week-end. Après, ces articles-là ou ces ébauches d'articles peuvent être utilisées comme point de départ pour un article de journaliste qui peut modifier l'article avec de la vraie recherche, la vraie valeur ajoutée du journaliste.

Relisez-vous les articles des matches de football générés automatiquement ?

Non, on ne les relit pas. Ici, on a validé le programme. Ça a été une très longue procédure. En premier lieu, on a dû marier quatre interlocuteurs, l'Union belge, LabSense, nous et des gens qui sont spécialisés dans les infographies et les tableaux de résultats, de la société IDalgo. Ça a été un processus très long de *mapping* des bases de données des uns et des autres. Il y a aussi un eu très gros travail de validation du texte, des formulations. C'est un peu du *fine-tunning*. Par exemple, ce n'est pas parce que le match se termine 4-1 que l'équipe battue a été écrasée. Parce qu'il peut y avoir des circonstances de jeu qui font que l'équipe n'a pas été complètement écrasée. L'intelligence du texte est réglée une fois pour toutes. Et ensuite, la moulinette, elle tourne. Et à ce moment-là, on n'a pas les moyens de relire plusieurs milliers d'articles par semaine.

Ce sont des articles qui n'auraient donc pas pu être produits par des humains ?

Non, impossible. Et les journalistes ne doivent pas être inquiets par rapport à cette technologie. Il y a une phrase que j'aime bien, qui dit que si un journaliste pense qu'un robot peut lui prendre son travail, il a vraiment de quoi s'inquiéter et c'est peut-être qu'il a raison. Parce que ça veut dire que le journaliste pense lui-même qu'il n'a pas de valeur ajoutée. Moi, je pars vraiment du principe que les générateurs de textes, c'est des tableaux Excel. Ce n'est pas plus séduisant qu'un tableau Excel. L'intelligence, elle est dans la qualité du texte, dans la sémantique, mais aussi non c'est de la *data*, de la pure *data*. Et non, ça ne va pas voler le boulot des journalistes. Aujourd'hui, il y a encore des journalistes qui passent leur temps avec leurs petites mains le dimanche soir à taper dans des petites cases le nom de l'arbitre, qui a joué. Ça, c'est du travail inutile. Un journaliste a bien mieux à faire que d'aller encoder de la *data*.

C'est donc plus une opportunité qu'une menace pour le journalisme ?

Dans ce cas-ci, l'Union belge est très contente. L'Union belge, au-delà de sa vocation des diables rouges, elle veut faire progresser le foot en Belgique. Donc ils sont très contents qu'on couvre, même de façon automatisée, des compétitions qui n'étaient jusque-là pas du tout couvertes. On explore vraiment un nouvel espace, ça n'a pas changé du tout la manière dont les journées se passent pour nos journalistes. Ça a juste ajouté une couche, en fait.

Êtes-vous contents des résultats ?

Oui, on est contents. On est un groupe de presse régional donc on a vocation de faire vivre les communautés locales. En soi, ces articles-là, on ne peut pas dire qu'ils performant, qu'ils font des dizaines de milliers de vues. De toute façon, nous, des milliers de vues, on en a déjà largement assez. Sur notre site *SudInfo* par exemple, on dépasse le million de visiteurs uniques par jour. Donc ce qu'on cherche plutôt à faire avec LabSense, c'est de commencer à faire vivre une communauté autour de ces articles-là, de s'adresser aux entraîneurs, aux patrons de clubs sportifs et à créer un microcosme autour de ces articles. C'est aussi une manière de combattre les réseaux sociaux qui ont capté la conversation publique. Donc c'est ça qu'on est occupés à faire pour la saison prochaine. Autour de ces articles, on veut créer des communautés, créer un contexte, pourquoi pas des informations servicielles pour les entraîneurs qui veulent se perfectionner, ou comment acheter des ballons meilleur marché. Donc c'est ça qu'on cherche à faire. C'est rester l'endroit où on cause dans nos communautés locales.

Utilisez-vous LabSense dans d'autres domaines que le sport ?

Pour l'instant, on l'utilise pour le foot des jeunes et des amateurs. Ce serait assez facile de monter sur d'autres sports, mais on va plutôt chercher à consolider ce qu'on fait et ce microcosme. Mais par contre, on lance alors des articles météo, sur la météo locale. On va faire un article météo par localité automatiquement. On est bien d'accord qu'entre Liège et Herstal, la météo n'est pas très différente. À partir de cette semaine ou la semaine prochaine, on générera un article météo par localité et par jour. Il y a d'autres choses qu'on va continuer d'explorer, comme les indicateurs. Dans le cadre des élections locales communales de 2018, on générerait des textes sur des indicateurs tels que le prix des maisons, le niveau de vie moyen, le nombre de chômeurs. Et donc il y avait une page par commune avec des informations de type politique qui étaient écrites par des journalistes. Mais on avait aussi des informations plus contextuelles sur la commune, une sorte de portrait automatisé de la commune, qui était à 100 % basé sur des indicateurs qu'on trouve sur des sites de la région wallonne. C'est de la pure *data* transformée en texte. Là, on va essayer de faire un ou deux articles par commune et par semaine, qui seront un peu plus intelligents. Parce que là c'était fort statique comme articles. Et pour les articles qu'on compte faire, on aimerait croiser des indicateurs entre eux, pour que ça devienne plus journalistique. Parce que dire uniquement : « cette commune a une vie supérieure par rapport aux communes autour d'elles », ce n'est pas assez, il faut croiser ça avec d'autres indicateurs. Donc on va continuer à creuser ça. Et puis après, ce que nous

sommes occupés à regarder aussi, c'est comment variabiliser les articles pour parfois un peu tromper Google, en particulier pour éviter le « *duplicate content* ».

Annexe 6 : Interview de David Coppi **Journaliste politique au *Soir***

Entretien réalisé le 31 mars 2020

- Retranscription -

Que pensez-vous des générateurs automatiques de textes ?

Je pense que tous ces systèmes peuvent venir en soutien pour compiler de l'information, mais il y a deux problèmes, au départ et au bout du compte. Au départ, il y a quand même un doute à avoir en permanence sur la source de l'information. Ça, on ne pourra pas l'éviter : les sources de l'information sont-elles fiables ou pas ? Et qui peut nous garantir qu'elles sont fiables ? La meilleure source d'information, c'est quand vous, vous voyez quelque chose. Et encore, elles sont tout de même déformées sous vos yeux par la lumière, etc. Mais c'est la seule source d'information véritablement fiable. Donc quand on fait confiance à des modèles informatisés etc., au départ, est-ce que c'est fiable ? On avance, mais il peut y avoir de gros problèmes parce qu'il n'y a aucune certification, c'est piratable facilement. Et puis, au bout du compte, la différence, c'est que le journaliste apporte sa dimension humaine et toute sa subjectivité dans la relecture de l'information et dans le traitement, dans la restitution. Or, je pense qu'on s'est fait des films sur la capacité des machines. La machine est en place, mais en définitive, si on enlève la subjectivité du journaliste dans le récit et dans l'écriture, je pense qu'on enlève le journalisme. Donc je pense que les machines n'ont aucun intérêt.

Pour vous, un article doit-il contenir « de l'humain » ?

Je pense qu'on peut le retirer, mais alors on est dans un autre monde. C'est tout autre chose. Et on peut opter pour un autre monde, mais il n'y aura pas d'information réelle. La seule information véritablement valable au bout du compte, c'est quand quelqu'un vous raconte quelque chose avec sa subjectivité à lui. Le reste, c'est très intéressant, ça permet de compiler des informations, mais si on retire le dernier acte, on est dans autre chose, on n'est plus dans le journalisme.

Ces systèmes d'IA représentent-ils une menace pour le journalisme ?

Non, si c'est utilisé à bon escient ça peut aider à compiler des données statistiques, historiques etc. Tout ça, ça peut être fort utile, mais avec toujours le doute sur les sources. Moi, j'ai un gros problème avec les sources sur le net. Parce que c'est totalement invérifiable et je n'ai aucune confiance en ça. Moi, je l'utilise tout le temps pour le journal, mais on doit quand

même toujours garder cette distance critique. La subjectivité du journaliste peut être remplacée matériellement, mais en réalité, on perd le sens du journalisme, simplement. On doit garder une distance critique. La plus-value du journaliste, c'est la subjectivité, ce qui est dans vos mains, dans vos yeux et dans votre façon d'écrire les choses, plus ou moins bien, de travers ou pas. Et normalement, ça, c'est intouchable, ou alors on sort du journalisme. On peut sortir du journalisme, c'est un choix.

Annexe 7 : Interview de Julien Covolo Journaliste web à la RTBF

Entretien réalisé le 31 mars 2020

- Retranscription -

Intelligence artificielle et journalisme vont-ils de pair ?

Je pense que oui. Quand je vois tous les progrès faits en datajournalisme en quelques années, oui, je pense qu'il faudra s'en servir de toute façon, à terme. Au niveau des générateurs automatiques de textes, je trouve que c'est une bonne chose si ça permet de décharger le journaliste d'une partie du travail qui pourrait être jugé comme ingrat. Je pense notamment à ce qu'on appelle le « bâtonnage » de dépêches, donc le fait de copier-coller des dépêches, de mettre des images. J'ai fait ça pendant un an, je n'ai pas l'impression de faire du journalisme quand je fais ça. Tu ne fais que de la mise en page, tu ne cherches pas d'informations et tu n'expliques rien. Mais il faut absolument le faire parce que ça permet d'éviter d'envoyer une équipe à chaque fois qu'il se passe quelque chose et d'utiliser le travail des agences. À la RTBF en tout cas, c'est comme ça qu'on fonctionne. Le fait d'enrichir ces dépêches permet déjà d'apporter un plus, plutôt que de bêtement copier-coller le texte qu'on reçoit. Donc je pense que pour les élections, par exemple, oui, les générateurs de textes, c'est pratique. Comme ça, on n'a pas besoin de créer un article pour chaque commune. Maintenant, ce qui me fait peur, moi, c'est que ce soit utilisé pour remplacer un journaliste. Je crains notamment que dans les grands groupes de presse on en arrive à cette situation.

Cela vous fait-il peur ?

Oui, ça peut faire peur. Mais ce n'est pas la technologie en elle-même qui me fait peur, c'est ce que les gens pourraient en faire. Quelqu'un pourrait se dire qu'il n'a plus besoin de payer un journaliste pour écrire vu que c'est tout à fait factuel. Mais par contre, si c'est utilisé pour décharger le journaliste d'une partie de son travail et qu'il puisse, de son côté, faire de l'investigation et un travail qu'aucune machine ne pourra jamais faire, alors là, c'est autre chose. Le problème, c'est le regard qu'on a par rapport au travail du journaliste. Il y a des gens qui sont persuadés qu'un journaliste est juste une personne qui est payée pour poser des questions à d'autres, alors que non, il y a tout un travail d'enquête. Et je pense qu'une machine ne pourra jamais le faire. Je ne vois pas une machine aller trouver Stéphane Moreau pour lui demander de rendre des comptes, par exemple.

Si la technologie est utilisée à bon escient, peut-elle tout de même représenter une opportunité pour les journalistes ?

Dans les rédactions, oui. Quand je vois la quantité d'informations qu'il y a, une personne ne saurait pas parler de tout. Et le travail que font les agences de presse nous aide déjà vraiment beaucoup à ce niveau-là puisque tu n'as plus qu'à copier le texte et le mettre en page. Et toi, tu peux te concentrer sur un autre sujet, rédiger un article plus approfondi sur un sujet que tu juges plus important. Donc ça nous aide déjà pas mal et je pense que c'est même, à mon avis, du côté des agences de presse que ça va beaucoup plus servir parce qu'une dépêche d'agence peut se faire rédiger par une machine.

Seriez-vous d'accord d'enrichir et de corriger un article réalisé par un robot ?

Oui parce que c'est déjà un peu ce que j'ai l'impression de faire maintenant avec des dépêches Belga. On ne reçoit que le texte avec un titre, on ne sait pas qui a écrit ça, il y a des fautes qu'on doit corriger. Ce n'est jamais qu'un tableau avec des cases et du texte qu'on doit remettre en forme. Donc c'est déjà un peu ce qu'on fait maintenant, sauf que ça a été écrit par un être humain.

Selon vous, quelle est la plus-value du journaliste par rapport aux robots ?

C'est la subjectivité. On dit toujours que le journaliste doit être objectif, mais j'ai toujours été convaincu que non, dans le sens où rien que le choix du sujet, la manière d'en parler, la façon de hiérarchiser l'information, c'est subjectif. Il faut que ce soit un être humain qui fasse ce travail d'enquête, de tri dans les informations. Ça, c'est le boulot du journaliste, de trier toutes les informations qu'il reçoit.

Le style du journaliste est-il important ?

Oui, bien sûr, c'est très important. Mais ma réflexion va dans les deux sens. Des journalistes qui écrivent pour des agences ont déjà un ton beaucoup plus neutre. Et dans l'autre sens, ça ne m'étonnerait pas qu'un jour on me montre un programme d'intelligence artificielle qui arrive à écrire dans un style où on jugerait que c'est un humain qui l'a écrit, alors que non. Je ne sous-estime pas la capacité des machines, mais au niveau du travail qu'il y a derrière, on ne remplacera pas le journaliste. Et puis même dans la rédaction, ce genre de chose ne pourrait pas s'appliquer à tous les cas de figure. Maintenant, le problème, et c'est ça ma crainte, ce serait qu'on dise qu'à la place d'engager dix journalistes, on en engage huit et deux robots.

Annexe 8 : Interview de Nicola Becquet

Responsable de la stratégie numérique de *L'Écho*

Entretien réalisé le 2 avril 2020

- Retranscription -

**Vous êtes en train de développer un projet d'articles automatisés au sein de *Mediafin*.
Pouvez-vous m'en dire plus ?**

Oui, nous avons un projet d'automatisation de production de textes à partir de données chiffrées. Quotidiennement, on a un *Live Blog*, c'est-à-dire un article suivi qui donne l'actualité des marchés par petit poste tout au long de la journée. Donc il y a deux journalistes qui se relaient chaque jour à cette tâche. À partir de ce moment-là, on fait des posts sur l'état des marchés le matin, à l'ouverture des différentes bourses, tout au long de la journée, des bourses européennes jusqu'à l'ouverture de la bourse et la clôture de la bourse à Wall Street. Ça, c'est une partie du travail qui est fastidieuse puisqu'il s'agit de réécrire chaque jour à la même heure un texte similaire, ou qui varie évidemment en fonction des chiffres, mais qui est assez répétitif et qui demande d'aller chercher des données sur plusieurs sites et plateformes différentes. Ça a une vraie plus-value pour les lecteurs, mais du point de vue du journaliste, c'est fastidieux et répétitif. À partir de ce moment-là, ayant suivi ce que faisait l'entreprise Syllabs et d'autres entreprises dans le monde qui proposaient l'automatisation de production de textes à partir de données, je me suis dit qu'il y avait peut-être quelque chose à faire de ce côté-là. Donc j'ai essayé de confronter cette intuition avec les journalistes eux-mêmes, évidemment. La réponse a été : pourquoi pas, oui, ça peut être intéressant. On a creusé et on a déposé un dossier chez Google au fond pour l'innovation, qui s'appelle DNI, pour recevoir une aide financière pour développer ce robot, si on peut dire, ce calibrage qui permet de produire, à partir de données qui sont fournies par une société extérieure, les données en continu sur notre site pour tout ce qui est chiffres relatifs à la bourse. Et donc on a construit des *templates* de textes en fonction des variations des données chiffrées, avec des cas : le BEL20 ouvre en hausse, en baisse, stagne à tel point, telle et telle valeur le tire vers le haut, vers le bas ou le font chuter, etc. Ça, c'est le principe général. Et l'objectif derrière, c'est d'automatiser une tâche assez fastidieuse et ingrate pour les journalistes, de fournir des textes d'ouverture, de mi-séance et de clôture très rapidement sur le site au lecteur, encore plus rapidement que ce qu'on peut faire de façon humaine. Parce que le temps de collecter, de rédiger, de publier, ça prend un certain temps, alors que là, c'est automatique. Ça prendra

à peu près une minute après avoir reçu les données. Ça pourrait être encore plus rapide, mais on va d'abord valider dans un premier temps les posts produits par le robot. Et donc l'objectif final, c'est de libérer les journalistes de cette tâche pour qu'ils puissent se concentrer sur l'analyse, par exemple, pourquoi telle bourse ouvre en hausse, en baisse, et que les journalistes soient plus rapidement sur la plus-value journalistique et non pas sur le *reporting* de chiffres.

Les articles générés automatiquement seront-ils publiés tels quels ou seront-ils enrichis d'une analyse ?

Les articles seront publiés tels quels. Dans un premier temps, ils seront validés pour voir si tout fonctionne bien, mais publiés tels quels puisqu'ils sont rédigés en paragraphes, on a évidemment fait un effort sur la formulation. Et ensuite, le journaliste peut publier un autre post où il va analyser des choses et ainsi de suite, mais il a bien évidemment une liberté d'action par rapport à ça.

Étant donné que vous avez travaillé sur les formulations, peut-on dire que les articles posséderont votre patte journalistique ?

Oui, après elles sont structurées, lisibles, etc., mais ce n'est pas de la grande littérature non plus. L'objectif, c'est que ce ne soit pas juste des chiffres qui soient mis dans un tableau. Ça, c'est l'autre spécificité, c'est qu'il y a une partie textuelle, et selon les types d'articles pendant la journée, il y a des tableaux qui sont ajoutés, des *tops* et des *flops* ou des *tops* des cinq valeurs qui progressent, ou des bilans trimestriels, etc. Donc il y a une grande variété. Il y a treize posts par jour tout au long de la journée et il y a des posts hebdomadaires, trimestriels et annuels. De base, ces articles-là, c'est une demande de notre audience et c'est un format qui fonctionne très bien. Et parallèlement, on va faire autre chose. Il y a une autre partie du projet. Sur notre site, chaque internaute peut avoir un portefeuille d'actions virtuelles, donc il choisit ses propres valeurs comme s'il investissait dans un vrai portefeuille. On l'a mis en place et ce n'est toujours pas lancé pour l'instant, mais on a mis en place des bilans personnalisés des portefeuilles. Donc il y a 130 000 portefeuilles qui sont sur le site et on voudrait pouvoir envoyer chaque matin un bilan personnalisé : « Bonjour Monsieur/Madame untel, votre portefeuille évolue à la hausse, à la baisse, tiré par telle valeur ou telle valeur, etc. », plutôt que d'offrir seulement une sorte de tableau de bord avec uniquement des chiffres. Et ensuite, on verra si on peut envoyer des newsletters personnalisées, etc. Et donc l'ensemble de ces contenus sera dispersé sur le site aussi, pas uniquement dans le *Live Blog*,

mais dans les tableaux de cours, dans toute une série de pages, par exemple la page de référence du BEL20 ou du CAC40, etc. Et bien on pourra aller les reprendre dans des carrousels pour les mettre en valeur dans la journée.

Ce projet n'a jamais été réalisé en Belgique ?

Non. Même en Europe, c'est un projet assez spécifique. Il y a deux médias qui l'ont fait au Danemark et en Suède, mais d'une autre façon.

De quelle manière les journalistes ont-ils réagi ?

Les journalistes étaient assez enthousiastes dès le départ, ils ont construit des choses. C'est eux qui ont fait les *templates*, qui ont vérifié les textes, etc. Sur la fin du projet, c'était un peu plus compliqué parce que leur service a été un peu diminué, le nombre de personnes dans le service a été réduit pour différentes raisons, des maladies ou des indisponibilités. Et donc ça a mis, au dernier moment, un frein au projet parce qu'ils n'étaient plus disponibles pour suivre le projet, ils devaient déjà faire l'actu' tout court. Donc vers la fin, on a pris du retard sur la finalisation. Mais en tout cas, ça a été fait avec l'ensemble des journalistes, ils ont vraiment été acteurs dans le choix de le faire et dans la réalisation. L'accueil a été plutôt favorable. Parfois, il y a eu quelques incompréhensions de la part des journalistes sur la difficulté technique à réaliser ce type de choses. Eux écrivent un texte et ils se disent qu'il suffit de faire un texte à trous. C'est beaucoup plus complexe et puis il y a beaucoup de garde-fous sur la vérification des données, si tout est là, si tout est fonctionnel, tous les aspects de calendrier, avec les fermetures, les jours fériés, les exceptions, il y a quand même pas mal d'exceptions. Donc parfois il y a eu de l'incompréhension de la part des journalistes qui vont dire : « c'est aussi compliqué que ça ? », « pourquoi ça prend autant de temps ? », etc. La difficulté qu'il y a eu aussi, c'est que Syllabs, l'entreprise avec laquelle nous collaborons, est à Paris. Nous, nous sommes à Bruxelles. On aurait dû davantage se voir et se rencontrer. C'est vrai que le travail à distance était quand même des plus complexes pour être sûr que le message passe correctement.

Les journalistes ne sont pas sentis menacés par l'arrivée d'un robot ?

Non, au-delà des petites blagues du style « on va être remplacés », eux, ils voient bien la plus-value et la possibilité pour eux de se concentrer sur autre chose qui est plus intéressant. C'est une expérimentation qui vient de la rédaction et pas d'un service commercial ou quoi

que ce soit. Le bémol, c'est la fin du projet malheureusement, qui a été un peu retardé à cause de problèmes d'effectifs, mais qui n'ont rien n' à voir avec le projet en lui-même.

Quand le projet sera-t-il mené à terme ?

Normalement c'était juillet, mais avec le coronavirus, ça va prendre un peu de retard. Il y a encore une dernière ligne droite. Et la spécificité aussi, c'est qu'on le fait aussi en néerlandais. Et il y a aussi des spécificités dues à la langue et à l'organisation du côté néerlandophone du *Tijd*.

Allez-vous préciser que les articles sont réalisés par des robots ?

Oui, tout à fait. Sous chaque post, il y aura la phrase : « produit par Quotebot ». Et il y aura un lien vers un article qui existe, qui décrit d'où viennent les données, qui est derrière, etc. On assume totalement cette partie-là et c'est normal d'avoir une transparence par rapport au lecteur.

Au niveau de la déontologie journalistique, les générateurs automatiques de textes sont souvent critiqués. Qu'en pensez-vous ?

Est-ce qu'on peut parler d'éthique quand on produit des petits bulletins d'information automatisés à partir de chiffres ? Oui et non. Mais ça a été fait par des journalistes derrière, donc la partie éthique, déontologique, etc. a été filtrée par des journalistes eux-mêmes, qui ont rédigé les modèles. On les entourait de précautions pour préciser d'où viennent les données, comment elles sont traitées, par qui, etc. C'est une forme de mise en ligne et de rédaction spécifique pour mettre en forme des chiffres. Il n'y a pas de jugement, il y a des appréciations qui sont très techniques, voire scientifiques. Si le cours ou l'indice fluctue entre +0,1 et -0,1, on va dire qu'il stagne, c'était des réflexes et des pratiques déjà utilisées.

Quelle est la plus-value des journalistes par rapport à *Quotebot* ?

Tout ce qui est analyse, connaissance du contexte, style aussi. Ce qu'on souhaite, c'est mettre en valeur la plus-value journalistique.

Si le projet se déroule bien, pensez-vous élargir la technique à d'autres domaines ?

C'est une possibilité. J'aimerais d'abord finaliser ce projet-ci, mais ça pourrait s'élargir dans certains domaines, comme l'immobilier, des choses comme ça.

Annexe 9 : Interview de Claude de Loupy **CEO de la société Syllabs**

Entretien réalisé par e-mail le 9 avril 2020

- Retranscription -

Parmi vos clients, comptez-vous beaucoup de médias ?

Les médias sont notre principal secteur d'activité. Ils représentent plus de 50 % de notre chiffre d'affaires en 2019. Au total, nous avons travaillé avec plus de 40 marques médias.

Combien de temps cela vous-a-t-il pris pour arriver au résultat de génération de textes que vous proposez aujourd'hui ?

La première version du moteur de rédaction a produit ses premiers textes pour des clients en 2011 en se basant sur cinq ans de développements préalables dans le domaine du traitement des langues.

Êtes-vous toujours en train d'améliorer les logiciels ?

Nous continuons depuis neuf ans à faire évoluer le système. De nombreuses fonctionnalités sont ajoutées au fur et à mesure. Nous travaillons également sur les systèmes annexes : production de graphiques pour accompagner les textes, base de connaissances pour enrichir les textes avec des informations croisées, développement de nouvelles langues, plateforme de diffusion des textes produits et, bien sûr, amélioration des capacités linguistiques.

Que pensez-vous pouvoir apporter aux médias aujourd'hui ?

Pour un média donné :

- Capacité à gérer la "longue traîne" en produisant des contenus qui intéressent un faible nombre de lecteurs mais dont la somme représente beaucoup de nouveaux lecteurs qui n'étaient pas adressés auparavant.
- Augmentation de l'empreinte du média sur le web par la mise à disposition d'un nombre de contenus plus importants. Cela implique le référencement dans les moteurs de recherche sur des requêtes où le média n'était pas présent (résultats élections municipales + nom d'un village) et le renforcement de positions sur des requêtes plus généralistes (résultats élections municipales).

- Libération de temps pour les journalistes sur des tâches que personne n'a envie de faire et où la capacité de réflexion et d'écriture des journalistes n'est pas sollicitée (météo par exemple).
- Capacité à traiter des informations en temps réel, quitte à ce qu'un journaliste enrichisse cette information par la suite.
- Une source d'information en cas de non-disponibilité imprévue (la personne devant couvrir tel match est malade, personne pour la remplacer au dernier moment, les contenus Syllabs peuvent servir de remplacement même si moins qualitatif).

Pour les médias locaux, on a en plus :

- La capacité à alimenter régulièrement ses lecteurs d'informations hyperlocales tout en conservant la rentabilité.
- La capacité à gérer de manière plus simple les correspondant locaux pour certaines informations qui peuvent être traduites sous forme de formulaires (compte-rendu de rencontre sportive...).
- Les médias locaux ont une importance considérable pour la préservation de nos modèles de démocratie, comme le prouvent les conséquences de la fermeture d'un média local (par exemple aux USA). La solution que nous proposons est une aide (parmi d'autres), pour permettre aux médias locaux de continuer d'alimenter leurs pages locales tout en laissant les journalistes apporter une information plus fouillée ou plus humaine.

De manière générale, les entreprises médiatiques réagissent-elles de façon positive lorsque vous proposez vos services ?

Oui. Les médias voient généralement d'un bon œil tout ce que nos services peuvent leur apporter.

Les journalistes ont-ils (toujours) peur de se faire "remplacer" ?

Le déterminant n'est pas judicieux. Il y a toujours eu DES journalistes qui ont peur de se faire remplacer. C'est encore le cas aujourd'hui même si beaucoup de journalistes constatent qu'il n'y a pas eu de licenciement lié à l'utilisation des moteurs de rédaction et que cela est bénéfique pour les médias, les journalistes qui n'ont plus à faire certaines tâches ingrates et les lecteurs qui trouvent une information qu'ils ne trouvaient pas auparavant et qui les concerne de près. La peur engendrée par ces technologies est logique : toute nouvelle

technologie a généré des inquiétudes par rapport aux métiers qui ne l'utilisaient pas auparavant. L'intelligence Artificielle (terme qui ne correspond en rien à la réalité) est utilisée depuis un siècle par les romanciers et scénaristes pour créer des oeuvres dystopiques où la machine cherche à dominer les humains.

Aujourd'hui, les générateurs de textes en journalisme sont surtout utilisés dans le domaine du sport, des élections et de la finance. Aimerez-vous élargir cela à d'autres domaines ?

Nos clients médias utilisent nos moteurs de rédactions dans ces domaines mais également : la météo, des agendas culturels, des descriptifs de lieux, des descriptifs de produits, des informations pratiques, la pollution, les tremblements de terre... Et il y a beaucoup d'autres choses possibles. Nous travaillons avec d'autres secteurs qui nous utilisent pour produire des annonces (immobilières...), des rapports, des mails personnalisés...

Votre logiciel peut-il être créatif ?

Non. La notion de créativité est toute relative. Nous ne pouvons pas prédire ce que le moteur va écrire. Mais nous savons ce qu'il ne peut pas écrire car nous connaissons tous les paramètres qui sont utilisés. Le besoin auquel nous répondons exige de pouvoir garantir des textes parfaits et variables. Ces exigences ne permettent pas de laisser de la créativité au moteur. La créativité est possible dans une certaine mesure. Personne au monde n'a de système dont le niveau d'intelligence puisse être comparé à l'intelligence humaine. Les systèmes "artistes" (le producteur de tableau développé par Deep Mind) n'ont pas d'autre créativité que ce que ses créateurs lui ont donné.

Est-ce un but que vous voudriez atteindre ?

Actuellement, ce qui me semble atteignable et que j'aimerais pouvoir faire, c'est un système capable de créer des contes pour enfants de manière automatique. Il y a un côté créatif mais il restera contrôlé par les connaissances qui seront données au moteur.

Un journaliste aura-t-il toujours une plus-value par rapport à vos logiciels ?

"Toujours" est un terme compliqué à utiliser en science, mais j'aurais tendance à penser que oui. Supposons qu'une intelligence artificielle ayant un niveau suffisant pour faire un travail de journaliste soit créée un jour. Je ne vois rien dans les laboratoires du monde actuellement qui me permette de penser qu'on va y arriver dans 10, 20 ou même 1000 ans, mais imaginons.

Si une telle intelligence existe, elle ne fera que représenter un point de vue. Il est certainement impossible de pouvoir représenter toutes les individualités. Or, chaque individu apporte sa propre vision de monde. Je ne vois donc pas comment une machine pourrait supprimer la valeur ajoutée d'un individu de par le regard unique que celui-ci peut apporter.

Parmi vos clients du monde des médias, y a-t-il des entreprises belges ? Se montrent-elles intéressées ?

Nous travaillons avec Mediafin et avons travaillé par le passé avec le groupe Rossel.

Annexe 10 : Articles sélectionnés pour l'étude de contenu

Articles générés automatiquement :

<https://www.lameuse.be/519361/article/2020-02-16/provinciale-ii-serie-liege-succes-6-0-du-royal-sporting-club-beaufays-face-au>

<https://www.lameuse.be/519311/article/2020-02-16/provinciale-iv-serie-b-liege-succes-2-0-du-sc-oneutois-face-au-royal-fraiture>

<https://www.lameuse.be/508970/article/2020-01-26/u14-regionaux-superieurs-liege-succes-8-2-de-la-royale-alliance-melen-micheroux>

<https://www.lameuse.be/508327/article/2020-01-26/provinciale-ii-serie-liege-succes-5-1-du-royal-stade-waremmien-football-club>

<https://www.lameuse.be/508326/article/2020-01-26/provinciale-ii-serie-liege-succes-2-0-du-royal-football-club-hannutois-face-au>

<https://www.lameuse.be/508297/article/2020-01-26/provinciale-iv-serie-b-liege-succes-1-4-du-sc-oneutois-face-la-rcc-xhorisien>

<https://www.lameuse.be/508293/article/2020-01-26/reserve-provinciale-d-liege-succes-1-3-du-cs-bollandois-face-au-football-club>

<https://www.lameuse.be/504444/article/2020-01-19/provinciale-iv-serie-d-liege-succes-0-3-du-racing-ans-montegnee-football-club>

<https://www.lameuse.be/504437/article/2020-01-19/provinciale-ii-serie-b-liege-succes-2-3-du-football-club-warsage-face-las>

<https://www.lameuse.be/504413/article/2020-01-19/provinciale-iii-serie-b-liege-succes-2-3-du-royal-harze-football-club-face-au>

<https://www.lameuse.be/501034/article/2020-01-12/u14-regionaux-serie-c-liege-succes-2-1-du-rrc-stockay-warfusee-face-au-royal>

<https://www.lameuse.be/501026/article/2020-01-12/reserve-provinciale-e-liege-succes-1-2-de-lunion-roclenge-wonck-face-las>

<https://www.lameuse.be/500967/article/2020-01-12/2eme-amateurs-acff-nationale-succes-1-0-du-football-club-tilleur-face-au-res>

<https://www.lameuse.be/500351/article/2020-01-12/reserve-provinciale-h-liege-succes-4-0-du-racing-ans-montegnee-football-club>

<https://www.lameuse.be/493678/article/2019-12-22/provinciale-1-dames-liege-egalite-2-2-entre-le-sc-saive-et-le-royal-association>

<https://www.lameuse.be/493598/article/2019-12-22/provinciale-iii-serie-b-liege-egalite-1-1-entre-le-racing-ans-montegnee-football>

<https://www.lameuse.be/493570/article/2019-12-22/provinciale-ii-serie-b-liege-succes-2-3-du-football-club-warsage-face-au-mcs>

<https://www.lameuse.be/490401/article/2019-12-15/provinciale-ii-serie-b-liege-succes-0-2-du-sc-saive-face-au-football-club-hermee>

<https://www.lameuse.be/490273/article/2019-12-15/provinciale-iv-serie-b-liege-succes-4-0-de-la-royale-all-clavinoise-sc-face-au>

<https://www.lameuse.be/486579/article/2019-12-08/provinciale-iv-serie-f-liege-succes-1-0-de-lac-soiron-face-la-skill-racing-union>

<https://www.lameuse.be/486545/article/2019-12-08/provinciale-iii-serie-c-liege-succes-5-1-de-la-royale-etoile-sp-dalhemoise-face>

<https://www.lameuse.be/486453/article/2019-12-08/provinciale-iv-serie-d-liege-egalite-2-2-entre-le-cs-juprelle-et-lunion-roclenge>

<https://www.lameuse.be/483348/article/2019-12-01/u14-provinciaux-serie-b-liege-succes-5-2-du-football-club-herstal-face-au-res>

<https://www.lameuse.be/483285/article/2019-12-01/provinciale-iii-serie-b-liege-succes-2-1-du-racing-club-vaux-chaudfontaine-faces>

<https://www.lameuse.be/482836/article/2019-12-01/provinciale-iii-serie-c-liege-succes-2-3-du-royal-aubel-football-club-face-au-sc>

<https://www.lameuse.be/519124/article/2020-02-16/reserve-provinciale-d-liege-succes-3-0-du-racing-club-vaux-chaudfontaine-face-au>

<https://www.lameuse.be/508962/article/2020-01-26/u15-provinciaux-serie-liege-succes-2-1-du-royal-aywaille-football-club-face-au>

<https://www.lameuse.be/507872/article/2020-01-25/u21-provinciaux-liege-succes-5-3-de-la-js-pierreuse-face-au-royal-daring-club-de>

<https://www.lameuse.be/500553/article/2020-01-12/3eme-amateurs-acff-serie-b-nationale-succes-6-0-du-football-club-united-richelle>

<https://www.lameuse.be/487268/article/2019-12-08/u14-regionaux-superieurs-liege-succes-1-0-du-royal-herve-football-club-face-au>

<https://www.lameuse.be/519985/article/2020-02-16/provinciale-ii-serie-c-liege-succes-3-0-de-let-elsautoise-face-au-royal-football>

Articles rédigés par des journalistes :

<https://www.lameuse.be/523892/article/2020-02-23/las-eupen-prend-une-grosse-gifle-anderlecht>

<https://www.lameuse.be/523779/article/2020-02-23/des-regrets-et-de-la-frustration-charleroi>

<https://www.lameuse.be/521418/article/2020-02-19/meunier-nous-avons-un-peu-manque-de-volonte>

<https://www.lameuse.be/521415/article/2020-02-18/ligue-des-champions-haaland-met-le-psg-au-pied-du-mur-analyse>

<https://www.lameuse.be/521414/article/2020-02-18/ligue-des-champions-latletico-madrid-retrouve-son-ame>

<https://www.lameuse.be/519254/article/2020-02-16/le-bayern-munich-ecrase-cologne-et-reste-seul-en-tete-1-4>

<https://www.lameuse.be/518916/article/2020-02-15/anderlecht-se-saborde-malines>

<https://www.lameuse.be/507937/article/2020-01-25/sc-charleroi-kv-malines-le-brouillard-eu-raison-des-zebres>

<https://www.lameuse.be/507683/article/2020-01-24/ligue-1-rennes-arrache-le-nul-en-fin-de-match-nice-1-1>

<https://www.lameuse.be/507678/article/2020-01-24/d1a-face-ostende-le-standard-evite-de-peu-lhumiliation>

<https://www.lameuse.be/504667/article/2020-01-19/allemande-boyata-et-lukebakio-impuissants-face-au-bayern-munich>

<https://www.lameuse.be/491775/article/2019-12-17/dorian-dessoleil-decu-apres-charleroi-zulte-waregem-2-0>

<https://www.lameuse.be/491742/article/2019-12-17/standard-antwerp-un-duel-de-costauds>

<https://www.lameuse.be/491682/article/2019-12-17/coupe-de-belgique-charleroi-elimine-par-zulte-waregem-2-0>

<https://www.lameuse.be/487240/article/2019-12-08/ligue-1-marseille-renverse-bordeaux-et-repousse-lille-6-points>

TABLE DES FIGURES

Figure 1. Le processus de génération automatique d'articles de presse (Graefe, 2016, p.18).....	22
Figure 2. Ligne du temps des événements marquants de l'histoire du journalisme automatisé.	27
Figure 3. Exemple d'article généré automatiquement par Narrative Science sur le site de Forbes le 22/07/2015	29
Figure 4. Exemple d'article généré automatiquement par Automated Insights sur le site d'Associated Press le 8 novembre 2018.....	32
Figure 5. Exemple d'article généré automatiquement sur le site du LA Times le 28/03/2013.....	34
Figure 6. Exemple d'article généré automatiquement sur le site du Washington Post le 31/08/2017.....	35
Figure 7. Exemple de résumé d'article publié sur le site de Knowherenews.com le 9/04/2020.....	37
Figure 8. Exemple d'article généré automatiquement par Syllabs sur le site du monde.fr lors des élections départementales françaises de 2015.	40
Figure 9. Exemple de bulletin généré par Bxl'air bot sur le site www.bxlairbot.be	43
Figure 10. Exemple d'article généré par LabSense sur le site de Sudpresse le 8/03/2020..	44
Figure 11. Exemple de la photo accompagnant les résumés de matches de football générés automatiquement sur le site de Sudpresse.....	87
Figure 12. Exemple de la structure des trois premiers paragraphes d'un article rédigé automatiquement sur le site de La Meuse.....	90

TABLE DES MATIÈRES

REMERCIEMENTS	1
SOMMAIRE	2
GLOSSAIRE	3
INTRODUCTION	7
1. Mise en contexte	7
2. L’ascension des robots	8
3. Transformation, digitalisation et automatisation du monde des médias	9
4. Enjeux de l’utilisation de la génération automatique en presse écrite	11
5. Trois temps pour poser un regard sur l’IA en presse écrite	12
6. Méthodologie	14
7. Genèse et limites du mémoire	14
CHAPITRE 1 : HISTORIQUE PROBLÉMATISÉ	16
1.1. Avant-propos : âge de la singularité, IA faible et IA forte	16
1.2. Développement des générateurs automatiques de textes	18
1.2.1 Bref historique	18
1.2.2 Fonctionnement général	18
1.2.3 Une offre qui s’est diversifiée	19
1.3. Développement du journalisme automatisé	20
1.3.1. Une utilisation de la génération automatique de textes parmi d’autres	20
1.3.2. Premiers pas du journalisme automatisé	21
1.3.3. Fonctionnement général	21
1.3.4. Un travail de collaboration	23
1.3.5. Une tendance grandissante	24
1.4. Les États-Unis : berceau du robot-journalisme	28
1.4.1. Narrative science	28
1.4.2. Automated Insights	31
1.4.3. The Los Angeles Times	33
1.4.4. The Washington Post	34
1.4.5. Knowherenews	36
1.5. Robot-journalisme en Asie	38
1.5.1. Robot-journaliste en trois dimensions	38
1.5.2. Tencent	38
1.6. La France : un acteur éminent	39
1.6.1. Deux sociétés de génération automatisée de textes	39
1.6.2. Le Monde : une première mondiale	40
1.7. La Belgique : nouvel acteur	42
1.7.1. Première production automatisée d’informations en Belgique	42
1.7.2. Sudpresse	44
1.7.3. Mediafin	48
CHAPITRE 2 : QUESTIONS CLÉS ET ENJEUX	52

2.1. Mise en contexte	52
2.2. Éthique et déontologie.....	53
2.2.1. Code de déontologie des IA ?.....	53
2.2.2. La transparence.....	58
2.3. Devant la loi.....	60
2.4. Menaces du journalisme automatisé.....	63
2.4.1. Disparition du journalisme humain.....	63
2.4.2. Menaces pour la qualité du journalisme.....	69
2.5. Opportunités du journalisme automatisé	70
2.5.1. Nouveaux horizons créés par les limites de l'IA	70
2.5.2. Créativité de l'IA.....	74
2.5.3. Chiens de garde de la démocratie	77
2.5.4. Nouvelles habitudes de consommation	79
2.5.5. Objectivité.....	80
2.5.6. Opportunités pour les entreprises médiatiques	81
2.6. Réflexion	84
CHAPITRE 3 : ÉTUDE DE CONTENU.....	85
3.1. Présentation de l'étude de contenu	85
3.2. Observations générales et cadrage.....	86
3.3. Hiérarchisation des informations.....	89
3.4. Rhétorique et thèmes abordés.....	91
3.5. Fréquence des mots	92
3.6. Comparaison entre articles automatisés et articles traditionnels	93
3.7. Conclusion de l'analyse	95
CONCLUSION.....	96
BIBLIOGRAPHIE	101
Ouvrages et chapitres d'ouvrages	101
Articles scientifiques	102
Articles de presse	107
Sources universitaires	110
Sources audiovisuelles.....	110
Sources Internet.....	111
ANNEXES	114
Annexe 1 : Interview de Marie Guérin.....	114
Annexe 2 : Interview d'Émilie Mascia	117
Annexe 3 : Interview de Nadia Chater	120
Annexe 4 : Interview de Ricardo Gutiérrez	122
Annexe 5 : Interview d'André Thiel.....	126
Annexe 6 : Interview de David Coppi.....	130
Annexe 7 : Interview de Julien Covolo.....	132
Annexe 8 : Interview de Nicola Becquet.....	134

Annexe 9 : Interview de Claude de Loupy	138
Annexe 10 : Articles sélectionnés pour l'étude de contenu	142
TABLE DES FIGURES	145
TABLE DES MATIÈRES.....	146