

Analyse de l'impact des IA génératives sur la presse en ligne : anatomie d'un *newsbot* basé sur GPT

Robert Viseur, FWEG, Université de Mons, Belgique

Résumé :

L'année 2023 a été marquée par la diffusion rapide des *Large Language Models* tels que GPT utilisé par l'agent conversationnel ChatGPT. Une application de ces modèles a connu un essor plus discret : les *newsbots*. Si ces derniers ne sont pas nouveaux, la possibilité de générer un texte capable de rivaliser avec une production humaine leur a donné une nouvelle vitalité. Différentes limitations ont cependant émergé, notamment sur le plan des droits d'auteur et de la fiabilité des informations produites. Sur base de la plate-forme Génération Network et d'une preuve de concept inspirée par celle-ci, nous avons analysé le processus d'automatisation, ses limitations, puis discuté l'impact sur les modèles d'affaires des sites de presse en ligne. Nous mettons en évidence différents défis à relever sur les plans scientifiques (par exemple : évaluation de l'originalité et de la fiabilité des productions générées) et managériaux (par exemple : choix de positionnement, acquisition de nouvelles compétences et dépendance accrue aux plateformes).

Mots clés :

newsbot ; intelligence artificielle ; ChatGPT ; modèle d'affaires ; BMC.

Analysis of the impact of generative AI on the online press: anatomy of a GPT-based newsbot

Abstract :

The year 2023 saw the rapid spread of Large Language Models such as GPT, used by the ChatGPT conversational agent. One application of these models has seen a more discreet boom: newsbots. While newsbots are not new, the ability to generate text that can rival human output has given them a new lease of life. However, a number of limitations have emerged, particularly in terms of copyright and the reliability of the information produced. Based on the Génération Network platform and a proof of concept inspired by it, we analyse the automation process and its limitations, then discuss the impact on the business models of online press sites. We highlight various challenges to be investigated, both scientific (e.g. assessment of the originality and reliability of the productions generated) and managerial (e.g. choice of positioning, acquisition of new skills and increased dependence on platforms).

Keywords :

newsbot ; artificial intelligence ; ChatGPT ; business model ; BMC.

Analyse de l'impact des IA génératives sur la presse en ligne : anatomie d'un *newsbot* basé sur GPT

1. Introduction

Au second semestre 2023 une annonce secouait la presse en ligne dédiée aux nouvelles technologies. Ari Kouts, un consultant français en innovation employé par le société VISEO, présentait, d'abord sur Twitter (X) puis lors du [meetup](#) « *Generative AI Paris* » du 11 mai 2023, son expérimentation « *Génération Network* ». En pratique, cette appellation recouvre deux sites web, [Cuisine Generation](#) et [Tech Generation](#), dont la particularité est d'être totalement alimentés par des IA génératives. Allons-nous vers une substitution des journalistes par les IA génératives ?

Cette question, la presse spécialisée se l'est posée. Compte tenu de la concurrence potentielle exercée par des sites similaires à Tech Generation, lui aussi positionné sur les actualités technologiques, le site NextInpact s'inquiétait notamment des implications pour les sites de presse en ligne : « *Chômage technique pour la team Next Inpact ?* » (Saliou, 2023). Cette inquiétude faisait écho à la mise à disposition sur le site d'OpenAI, de [ChatGPT](#), un agent conversationnel généraliste, mais aussi du LLM (*Large Language Model*), baptisé GPT (Floridi & Chiriatti, 2020), sur lequel ce *chatbot* est construit. Pourtant, le potentiel de la NLG (*Natural Language Generation*), notamment dans le secteur de la presse, est un objet d'attention pour la recherche académique depuis les premières versions du modèle GPT (de-Lima-Santos & Ceron, 2020). Et, en pratique, le phénomène des *newsbots*, c'est-à-dire d'automates chargés de rédiger des actualités, n'est pas un phénomène inconnu.

Quel est au final l'impact potentiel de ces technologies sur le modèle d'affaires des sites d'actualités en ligne ? Pour répondre à cette question, notre recherche, exploratoire, est décomposée en quatre sections. Après une revue de la littérature dédiée aux IA génératives et aux *newsbots*, nous présentons notre méthodologie. Nous analysons ensuite le processus d'automatisation de sites d'actualités à l'aide de GPT puis discutons, avant de conclure, l'impact sur le modèle d'affaires des sites d'actualités.

2. Revue de la littérature

Cette section porte, d'une part, sur la présentation des intelligences artificielles génératives (IAG), en particulier celles du type LLM, d'autre part, sur le phénomène des *newsbots*.

2.1. Intelligences artificielles génératives

[ChatGPT](#), développé par OpenAI, est un agent conversationnel (*chatbot*) généraliste basé sur l'architecture GPT (*Generative Pre-trained Transformer*)¹. Il se distingue par sa faculté à interpréter des directives formulées en langage naturel, appelées « *prompts* », et à engendrer des réponses textuelles cohérentes en adéquation avec ces directives. GPT est « *un modèle*

1 Cf. <https://help.openai.com/en/articles/6783457-what-is-chatgpt>.

linguistique autorégressif de troisième génération qui utilise l'apprentissage profond pour produire des textes semblables à ceux des humains » (Floridi & Chiriatti, 2020). En tant que *Large Language Model* (LLM), il est capable de « *prédire statistiquement des séquences de mots* ». Ce modèle linguistique est formé par entraînement « *sur un ensemble de données non étiquetées composé de textes, tels que Wikipédia et de nombreux autres sites, principalement en anglais, mais aussi dans d'autres langues* ».

Les capacités de ChatGPT ont été améliorées au fil du temps, d'une part, grâce à l'amélioration du modèle GPT, d'autre part, suite à l'ajout de nouvelles fonctionnalités. ChatGPT est actuellement disponible en deux versions : ChatGPT (gratuit), basé sur GPT-3.5, et ChatGPT Plus (20 dollars / mois), donnant accès à GPT-4 ainsi qu'à des fonctionnalités complémentaires incluant l'accès au Web (navigation), des extensions tierces, l'analyse de données ou encore la production d'images avec [DALL-E](#). D'autres agents conversationnels (p. ex. [Google Bard](#)) et d'autres modèles de type LLM (p. ex. [Llama](#) et [Mistral](#)) font aujourd'hui concurrence aux produits d'OpenAI (Singh et al., 2023). À côté de l'accès au *chatbot* en mode SaaS (*Software as a Service*), OpenAI propose également un ensemble d'API permettant d'intégrer ses modèles de génération de texte au sein d'applications avec une tarification au *pay-per-use*².

Ces technologies ont entraîné un ensemble de craintes sur le plan éthique, liées notamment à la difficulté à distinguer les contenus générés de ceux écrits par un rédacteur humain, aux limitations des jeux de données utilisés ou encore à la fiabilité des informations publiées (Stahl & Eke, 2023). Concernant ce dernier point, ChatGPT n'a en effet pas « *de réelle compréhension des textes qu'il produit* ». Cela se traduit par le phénomène d'« *hallucination* »³ (également appelé *confabulation* ou *délusion*), soit la capacité des modèles IA de générer du contenu qui n'est pas basé sur des données du monde réel, mais plutôt sur l'imagination du modèle (Beutel, Geerits et Kielstein, 2023). Ce comportement entraîne un risque « *de mésinformation et de désinformation* » (Stahl & Eke, 2023), particulièrement dérangeante dans un contexte journalistique où la profession est supposée « *respecter la vérité* » et « *publier seulement des informations dont l'origine est connue* »⁴.

2.1. Phénomène des *newsbots*

L'entreprise [NewsGuard](#), spécialisée dans la lutte contre les *fake news*, a pourtant détecté plusieurs dizaines de *newsbots* en mai 2023 (McKenzie & Arvanitis, 2023). Elle soulignait en particulier les problèmes de qualité rédactionnelle causés par ces outils, incluant les messages d'erreur non traités avant publication, la publication d'informations erronées et les problèmes de contenus partiellement plagés, sans que les causes de ces dysfonctionnements ne soient clairement explicitées. S'agit-il en effet de limitations inhérentes à cette nouvelle technologie des IA génératives ou d'un défaut d'implémentation des *prompts* (c'est-à-dire la séquence d'instructions envoyée à l'IA générative à laquelle la rédaction est déléguée) ? Une partie au

2 Cf. <https://openai.com/blog/openai-api> pour plus d'informations à propos des API offertes par OpenAI.

3 Cf. <https://www.ibm.com/topics/ai-hallucinations>.

4 Cf. <https://www.ajp.be/telechargements/droitsdevoirs.pdf> (Charte de Munich, 1972).

moins du problème provenait, au vu des messages d'erreur non filtrés, d'un usage de l'IA générative sans information en entrée, soit un usage propice aux hallucinations.

Une démarche ambitieuse a également été mise en œuvre par la société cotée en bourse [CNET](#). Cette dernière publie des sites d'actualités technologiques et a discrètement profité des IA génératives émergentes pour produire des actualités de manière automatisée. L'impact sur l'image s'est cependant révélé négatif puisque des problèmes d'informations erronées et de plagiat au détriment de sites concurrents ont également été constatés (Christian, 2023). Là aussi, le manque de transparence sur les *prompts* utilisés ne permet pas de trancher sur les causes de ces dysfonctionnements d'une technologie par ailleurs exploitée dans ses premières versions. Que peut dès lors nous apprendre l'expérimentation « Tech Generation » sur le (dys)fonctionnement de ces *newsbots* dopés à l'IA générative ?

3. Méthodologie

Dans cette recherche exploratoire, deux ensembles distincts de données ont été exploités, liés à deux projets spécifiques : la plate-forme « *Tech Génération* » et une preuve de concept développée *ex nihilo*.

Le site [Tech Generation](#) est consultable en ligne. Il peut donc être analysé en tant que produit d'une technologie d'IA générative. De plus, la presse conventionnelle, à laquelle nous associons les *pure players* spécialisés, se sont penchés sur ce phénomène et n'ont pas manqué d'en critiquer le fonctionnement. Ari Kouts, son créateur, un consultant français en innovation employé par la société VISEO, a par ailleurs présenté son projet, d'abord sur Twitter (X) puis lors du [meetup](#) « *Generative AI Paris* » du 11 mai 2023.

Notre preuve de concept suit une architecture similaire à celle de Tech Generation. Il comporte un *newsbot* capable de collecter un ensemble de communiqués de presse, d'en conserver les plus intéressants puis de les transformer dans un style succinct en vue de leur publication sur un site web. Il nous permet en pratique de discuter les choix techniques envisageables et d'identifier les éléments posant des difficultés.

4. Processus d'automatisation

Le processus d'automatisation associé à un *newsbot* est analysé en deux étapes. Premièrement, nous analysons le processus mis en œuvre par Tech Generation. Deuxièmement, nous décrivons les choix techniques retenus pour notre *newsbot*, puis en analysons certaines limitations

4.1. Automatisation en quatre étapes

La plate-forme développée par [Ari Kouts](#) comporte quatre processus (Kouts, 2023 ; Saliou, 2023). Le premier processus collecte des actualités potentiellement intéressantes. Le modèle retenu consiste à identifier des actualités publiées aux États-Unis puis à les publier en France sous une forme réécrite, soit un modèle courant dans le monde des blogs d'actualités technologiques. La collecte peut par exemple être réalisée sur base des flux [RSS](#) de sites d'actualités faisant référence. Le second processus joue le rôle d'un rédacteur en chef et décide des sujets qui vont être traités. Ce processus est délégué à GPT. Le troisième processus

consiste à générer les actualités en français. La rédaction est précédée d'une étape de création de personas de rédacteurs permettant de conférer une personnalité à l'IA générative et de modifier le style de rédaction. Le quatrième processus inclut la publication des actualités. Le texte est complété par une image automatiquement récupérée dans une banque d'images dites « libres de droit ». En mai 2023, ce site d'actualités avait dépassé le millier d'actualités. Son référencement se révélait d'ailleurs un succès avec près de 1000 pages indexées dans Google Web Search, plus de 2000 pages dans Bing, fin octobre 2023. Google ne rejette en effet pas les contenus générés par intelligence artificielle tant qu'ils sont adaptés à une lecture humaine. Google Actualités a pour sa part accepté l'inclusion du site (Turcan, 2023).

4.2. Choix de conception

La rédaction des articles de presse peut être manuelle (rédacteur salarié *freelance* ; p. ex. [Fiverr](#)) ou automatique. Dans ce second cas, plusieurs sources peuvent être exploitées : API de moteur de recherche (p. ex. [Bing News API](#)) ou fichiers de syndication ([RSS](#), [Atom](#)) alimentés par des sites de la presse spécialisée (p. ex. [Techcrunch](#)) ou d'entreprises (blogs, communiqués de presse). En cas d'absence (fréquente) de fichier de syndication, le *scraping* des actualités à partir de l'HTML (métadonnées au format [JSON-LD](#) ou données associées aux balises) reste possible. Nous avons retenu l'utilisation de communiqués de presse, dans un premier temps via les fichiers de syndication.

Ces actualités peuvent ensuite être filtrées, soit par un rédacteur en chef humain, soit par un système d'apprentissage supervisé local, soit en utilisant un LLM. Nous avons fait le choix, comme Tech Generation, d'utiliser GPT-4. Ce dernier est utilisé pour catégoriser les articles puis les évaluer sur une échelle de 1 à 3. Le *prompt* précise le rôle affecté à l'IA, le public cible, les tâches et le format de réponse ([CSV](#)). Les communiqués de presse sélectionnés sont ensuite utilisés pour générer une brève avec l'API GPT-4. La réponse contient un titre, un corps d'article et différentes informations complémentaires incluant les personnes, les marques et les produits cités. Le *prompt*, préalablement testé dans ChatGPT Plus, précise le rôle affecté à l'IA, le public cible, les tâches, le format de réponse et les contraintes (fond, forme). Compte tenu de la limite de taille de contexte d'un *prompt*, les articles initiaux sont réduits, si nécessaire, à plus ou moins 5000 caractères à l'aide de GPT-3.5. La dépêche (titre, corps) est ensuite traduite avec l'API de DeepL⁵.

ChatGPT Plus a été utilisé pour l'aide à la programmation du *newsbot* (Python) mais aussi pour la création du site web (PHP) : création d'un logo (DALL-E), création d'une structure HTML (incluant les métadonnées LD-JSON), aide au codage, traduction des éléments de l'interface multilingue, structure de fichier RSS et structure de fichier « sitemap.xml ». Le site a fait l'objet d'un effort spécifique d'optimisation du référencement (Shahzad et al., 2020) incluant l'optimisation du contenu (texte et maillage), la réécriture des URL, l'insertion de mots-clés dans les URL ou encore l'individualisation des balises META. L'ensemble a nécessité moins de 5 jours de travail.

5 Cf. <https://www.deepl.com/fr/docs-api> pour plus d'informations. Voir les langues supportées par DeepL Pro : <https://support.deepl.com/hc/en-us/articles/360019925219-Languages-included-in-DeepL-Pro>.

4.3. Limitations liées aux LLM

Ces choix de conception nous permettent d'identifier certaines limitations à ce processus d'automatisation.

Premièrement, l'utilisation d'un LLM à des fins d'évaluation de communiqués de presse - une pratique également constatée pour l'évaluation multicritère sur base d'avis en ligne par exemple - pose question quant à sa fiabilité. Le LLM, en tant que générateur de texte, est capable de fournir une réponse dans le format imposé. Pour autant, les évaluations fournies font-elles sens ? Ce point mériterait une évaluation spécifique.

Deuxièmement, l'utilisation d'un LLM pour de la création d'articles amène rapidement la question de la contrôlabilité du résultat. Le style (longueur, tonalité...) est-il respecté ? Le texte est-il original ? Le résultat obtenu est-il reproductible ? L'utilisation d'un texte de départ protège-t-il du phénomène d'hallucination ? Les performances sont-elles stables dans le temps ? Notre expérimentation permet de constater que l'obtention systématique de textes totalement originaux, c'est-à-dire sans reprise à l'identique de séquences d'entrée, est relativement difficile. Il en va de même pour des contraintes strictes de taille. La littérature constate en effet que les LLM ne sont exempts, ni de régurgitations (phénomène de mémorisation ; de Wynter et al., 2023), ni surtout d'hallucinations, même en cas de tâches simples (p. ex. production d'un résumé basée sur un texte imposé ; Luo et al., 2024).

Ces constats amènent plusieurs perspectives de recherche. Comment caractériser, et surveiller, l'originalité d'un texte généré à partir d'un texte de départ ? Comment détecter une hallucination dans un texte généré ? Comment détecter une baisse de performance globale ?

5. Discussion

Cette section comporte une discussion des impacts des IA génératives sur le modèle d'affaires de la presse en ligne puis la proposition d'un continuum de degrés d'automatisation.

5.1. Impacts sur le modèle d'affaires

Ces nouvelles technologies amènent donc leur lot d'opportunités pour les éditeurs de sites spécialisés mais avec, d'une part, des limitations, d'autre part, une innovation de modèle d'affaires. Nous discutons ici la création de valeur et la captation de valeur (Osterwalder & Pigneur, 2011 ; cf. Tableau 2 *infra*). L'analyse des coûts s'appuiera sur le concept de coût total de possession (TCO) selon la définition donnée par Shaikh et Cornford (2011).

Premièrement, cette approche de la création de contenus suscite des espoirs en matière de réduction des coûts (cf. Tableau 1 *infra*). Là où la création d'un article sur une plate-forme de *freelances* type [Fiverr](#) coûte de l'ordre de 15 euros pour un article de 1000 mots, l'utilisation de l'API GPT-4-32K (soit la plus puissante ; voir la [grille tarifaire](#)) coûte moins de 10 cents pour 1000 tokens. La tarification est réalisée en *pay-per-use* ; elle se calcule à partir de la taille de la requête incluant la consigne et les données. Un bloc de 1000 tokens équivaut approximativement à 750 mots. La génération d'une dépêche coûte dès lors une fraction d'une création externalisée tout en étant plus rapide (compter de l'ordre d'une minute par *prompt*) et disponible 24h/24. Par ailleurs, l'arbitrage entre modèles, en particulier entre GPT-

3.5 et GPT-4, dépend non seulement du niveau de performances attendu mais aussi des contraintes financières ; les modèles moins puissants sont ainsi sensiblement moins coûteux.

Données :

		Prix	Mots
Fiverr	Rédaction	15,00 €	1.000

		In	Out	Tokens
OpenAI	GPT-4 8K	\$0,03	\$0,06	1000
	GPT-3-5 16K	\$0,003	\$0,004	1000

	Fixe	Pay per use	Caractères
Deepl	4,99 €	20,00 €	1.000.000

USD / EUR	0,95
Mots / Tokens	0,75
Caractères / mots	5

Objectif journalier	10
Langue UE	24
Fréquence	24
Taille (mots)	500

Prompt (caractères)	2500
---------------------	------

Simulation :

	Fiverr	Fiverr + Deepl	OpenAI	OpenAI + Deepl
Rédaction (en)	7,50 €	7,50 €	0,28 €	0,28 €
Rédaction (UE)	172,50 €		6,33 €	
Traduction		0,004 €		0,004 €
Coût par article	180,00 €	7,50 €	6,60 €	0,28 €
Coût total (jour)	1.800,00 €	75,04 €	66,03 €	2,79 €
Coût total (mois)	54.000,00 €	2.251,09 €	1.980,95 €	83,63 €

Tokens	750
Caractères	575.000

Tableau 1. Coûts associés aux modalités de production d'articles (site en 24 langues).

Deuxièmement, et contrairement peut-être à certains espoirs initiaux, la programmation d'une IA générative n'est pas simple. La compétence de création, et de mise à jour, des *prompts* implique la construction de compétences internes (activités clefs), et le cas échéant la recherche de compétences externes stables (partenaires clefs), en plus de leur stockage une fois mis au point (ressources clefs). La gestion des *prompts* nécessite donc de pouvoir investir. Elle peut d'ailleurs s'accompagner de coûts supplémentaires. D'une part, le modèle GPT-4 est un modèle LLM générique. Il est cependant possible de le spécialiser (p. ex. *fine tuning*), ce qui suppose davantage de compétences et de ressources matérielles. D'autre part, le recours à cette API entraîne une dépendance à OpenAI. Ces technologies d'IA générative pourrait donc venir renforcer la dépendance de la presse en ligne aux grandes plates-formes (Ouakrat 2020 ; Rebillard & Smyrniotis, 2010). Actuellement liée à l'apport d'audience (moteur de recherche ; p. ex. Google Actualités), aux revenus (régies publicitaires ; p. ex. Google Ads) et à l'analyse de données, cette dépendance pourrait dès lors s'étendre à la production des contenus eux-mêmes. Une gestion pro-active des risques associés peut donc plaider pour l'usage d'un modèle génératif local et (davantage) maîtrisé basé sur une technologie open-source (p. ex. [LLaMa](#)). Au delà de la réduction du *vendor lock-in*, l'utilisation de modèles exécutés localement devient rentable au-delà d'environ 30 articles générés quotidiennement.

Troisièmement, les IA génératives permettent de démultiplier à moindre coût la communication (canaux). Sur base d'un contenu validé, elles permettent en effet la rédaction de messages à destination des réseaux sociaux numériques. Sur base des [API X](#), par exemple, l'actualité peut être répercutée dès sa publication dans un format court et adapté. Ces comptes

constituent alors des sources de revenus indirects (modèle *freemium*) et directs dès lors qu'ils peuvent être monétisés⁶.

Quatrièmement, les IA génératives autorisent la diversification des contenus en vue d'atteindre des lectorats de niche voire d'individualiser les articles ou de proposer des revues de presse personnalisées. Plus coûteux, en particulier si la génération s'appuie sur une API, ce modèle suppose une tarification de la consultation des articles que la personnalisation permettrait de justifier.

Partenaires spécialisés en IAG pour le support au développement et à la maintenance des <i>prompts</i> .	Développement et maintenance des <i>prompts</i> (si capacités internes).	Proposition d'articles d'actualités thématiques et frais.	Rédaction d'articles personnalisés.	Rédaction automatisée d'articles pour des niches étroites de marché.
	Gestion des connaissances liées aux <i>prompts</i> .		Communication automatisée sur plusieurs canaux : Twitter (X), Facebook...	
Coûts de développement et de maintenance des <i>prompts</i> .		Revenus supplémentaires liés à la couverture de niches de marché.		
Coûts d'utilisation (<i>pay-per-use</i>) des API des IA génératives (GPT et Deepl).				
		Revenus supplémentaires liés à la personnalisation des articles (style, synthèse...).		
		Revenus supplémentaires issus des monétisations complémentaires (p. ex. compte X).		

Tableau 2. Modèle d'affaires (BMC) d'un site d'actualités co-généré.

Les performances croissantes des intelligences artificielles génératives permettent d'envisager différents degrés d'automatisation.

5.2. Degré d'automatisation

L'automatisation poursuit un processus en quatre étapes : artisanat, taylorisation, externalisation puis automatisation (de certaines tâches préalablement taylorisées). Nous nous concentrons ici sur les modes d'automatisation liés à l'écriture des articles publiés. Le premier est l'automatisation complète. Dans ce schéma, les articles sont générés avec un LLM tel que GPT-4 dans une langue de référence (p. ex. anglais) puis démultipliés à moindre coût à l'aide de Deepl. Nous qualifions ce modèle d'« *information ultra low cost* ». Le second est l'automatisation complète individualisée. Les articles sont alors générés par un LLM pour chaque langue, ce qui permet une adaptation à la culture ou aux actualités locales. Nous qualifions ce modèle d'« *information low cost* ». Une approche moins radicale de l'automatisation consiste à décrire la trame d'un article pour ensuite laisser un LLM développer l'article sur base de consignes de style. Nous qualifions ce modèle de « *journalisme augmenté* ». Le travail de rédaction de cette trame peut être confié à des journalistes associés, à des pigistes ou à des rédacteurs *freelances*, en fonction du degré d'externalisation de cette tâche. Les modèles qualifiés d'information *low cost* ou *ultra low*

6 Cf. <https://help.twitter.com/fr/rules-and-policies/content-monetization-standards> pour plus d'informations relatives à la monétisation des comptes Twitter (X).

cost reprennent en pratique un principe de création des actualités assez classique chez les blogs. Si certains blogueurs rédigent sur base d'informations qu'ils ont eux-mêmes collectées, d'autres s'appuient en effet sur les contenus publiés par les médias mainstreams (Wei, 2009). C'est d'ailleurs la stratégie adoptée par le site Tech Generation.

6. Conclusion

Au delà des désillusions ou des attentes irréalistes propres à toute *hype* technologique, les *newsbots* recèlent un risque de disruption par le bas pour les sites d'actualités basés sur la publication de dépêches sans réelle valeur ajoutée (ce que nous avons qualifié d'information *low cost* ou *ultra low cost*). Ils nécessitent cependant le développement de compétences et de ressources spécifiques pour atteindre et maintenir un niveau de qualité suffisant. Les technologies sous-jacentes ouvrent par ailleurs d'autres perspectives, notamment en matière de captation de revenus, avec par exemple la personnalisation des dépêches.

Cette recherche exploratoire alimente un agenda de recherche en matière de contrôlabilité des IA génératives dans le contexte du secteur de la presse et identifie des perspectives d'automatisation impactant le modèle d'affaires. Elle met par ailleurs en évidence la complexité liée à l'intégration de ces technologies. Dès lors, elle permet d'entrevoir les freins à la diffusion de ces technologies (Rogers, 2003). Comment les groupes de presse prennent-ils leur décision d'adopter ou non ces technologies ? Comment les ressources informatiques internes ou externes concourent-elles à cette diffusion ? Nous pensons par exemple à des entreprises de services numériques (ESN) telles [Audaxis](#), liée au groupe de presse [Rossel](#)⁷, actif sur la Belgique francophone (Le Soir, Sudinfo...) et les Hauts-de-France (Nordclair, Courrier picard...), ou à des initiatives liées aux « *Big Tech* » (p. ex. [Google News Initiative](#)).

7. Références

- Beutel, G., Geerits, E., & Kielstein, J. T. (2023). Artificial hallucination: GPT on LSD?. *Critical Care*, 27(1), 148. <https://doi.org/10.1186/s13054-023-04425-6>.
- Cardon, D. (2019). *Culture numérique*. Presses de Sciences Po. ISBN : 978-2724623659.
- Christian, J. (2023). CNET's AI Journalist Appears to Have Committed Extensive Plagiarism. *Futurism*, 23 janvier 2023. <https://futurism.com/cnet-ai-plagiarism>.
- de Wynter, A., Wang, X., Sokolov, A., Gu, Q., & Chen, S. Q. (2023). An evaluation on large language model outputs: Discourse and memorization. *Natural Language Processing Journal*, 4. <https://doi.org/10.1016/j.nlp.2023.100024>.
- de-Lima-Santos, M. F., & Ceron, W. (2021). Artificial intelligence in news media: current perceptions and future outlook. *Journalism and media*, 3(1), 13-26. <https://doi.org/10.3390/journalmedia3010002>.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds & Machines* 30, 681–694. <https://doi.org/10.1007/s11023-020-09548-1>.
- Kouts, A. (2023). On laisse les IA aux commandes ? Un PoC de sites autonomes. *Generative AI Paris*, 11 mai 2023. <https://www.youtube.com/watch?v=WDS8ozz8mH4>.

7 Cf. <http://www.actionnariatwallon.be/organisme/audaxis-20896>.

- Luo, J., Li, T., Wu, D., Jenkin, M., Liu, S., & Dudek, G. (2024). Hallucination Detection and Hallucination Mitigation: An Investigation. arXiv preprint arXiv:2401.08358. <https://doi.org/10.48550/arXiv.2401.08358>.
- McKenzie, S. & Arvanitis, L. (2023). Les “newsbots” montent au front : des sites d’actualité générés par l’IA se multiplient en ligne. NewsGuard, 01 mai 2023. <https://www.newsguardtech.com/fr/special-reports/bots-ia-generative-sites/>.
- Osterwalder, A., & Pigneur, Y. (2011). Business Model New Generation. Pearson. ISBN : 978-2744064876.
- Ouakrat, A. (2020). Négocier la dépendance? Google, la presse et le droit voisin. Sur le journalisme, 9(1), 44-57. <https://doi.org/10.25200/SLJ.v9.n1.2020.417>.
- Rebillard, F., & Smyrnaio, N. (2010). Les infomédiaires, au coeur de la filière de l’information en ligne: les cas de Google, Wikio et Paperblog. Réseaux, (2), 163-194. <https://doi.org/10.3917/res.160.0163>.
- Rogers, E.M. (2003). Diffusion of innovations - Fifth edition. The Free Press. ISBN : 0-7432-5823-1.
- Saliou, M. (2023). Tech Generation, le site qui teste ce que ChatGPT sait faire en journalisme tech. Next Impact, 4 mai 2023. <https://www.nextinpact.com/article/71594/tech-generation-site-qui-teste-ce-que-chatgpt-sait-faire-en-journalisme-tech>.
- Shahzad, A., Jacob, D. W., Nawi, N. M., Mahdin, H., & Saputri, M. E. (2020). The new trend for search engine optimization, tools and techniques. Indonesian Journal of Electrical Engineering and Computer Science, 18(3), 1568-1583. <https://pdfs.semanticscholar.org/350a/4f1fca7429c2c1cb3381f929fa2e9eb6df85.pdf>.
- Schiff, F. (2003). Business models of news Web sites: A survey of empirical trends and expert opinion. First Monday. <https://doi.org/10.5210/fm.v8i6.1061>.
- Shaikh, M., & Cornford, T. (2011). Total cost of ownership of open source software: a report for the UK Cabinet Office supported by OpenForum Europe. [https://eprints.lse.ac.uk/39826/1/Total_cost_of_ownership_of_open_source_software_\(LSER_O\).pdf](https://eprints.lse.ac.uk/39826/1/Total_cost_of_ownership_of_open_source_software_(LSER_O).pdf).
- Singh, S. K., Kumar, S., & Mehra, P. S. (2023). Chat GPT & Google Bard AI: A Review. In 2023 International Conference on IoT, Communication and Automation Technology (ICICAT) (pp. 1-6). IEEE. <https://doi.org/10.1109/ICICAT57735.2023.10263706>.
- Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT—Exploring the ethical issues of an emerging technology. International Journal of Information Management, 74, 102700. <https://doi.org/10.1016/j.ijinfomgt.2023.102700>.
- Turcan, M. (2023). Qui peut arrêter Tech-Generation, le site d’info tech généré par ChatGPT ? Next Impact, 9 décembre 2023. <https://www.numerama.com/tech/1584908-qui-peut-arreter-tech-generation-le-site-dinfo-tech-genere-par-ia.html>.
- Wei, L. (2009). Filter blogs vs. personal journals: Understanding the knowledge production gap on the Internet. Journal of computer-mediated communication, 14(3), 532-558. <https://doi.org/10.1111/j.1083-6101.2009.01452.x/>.