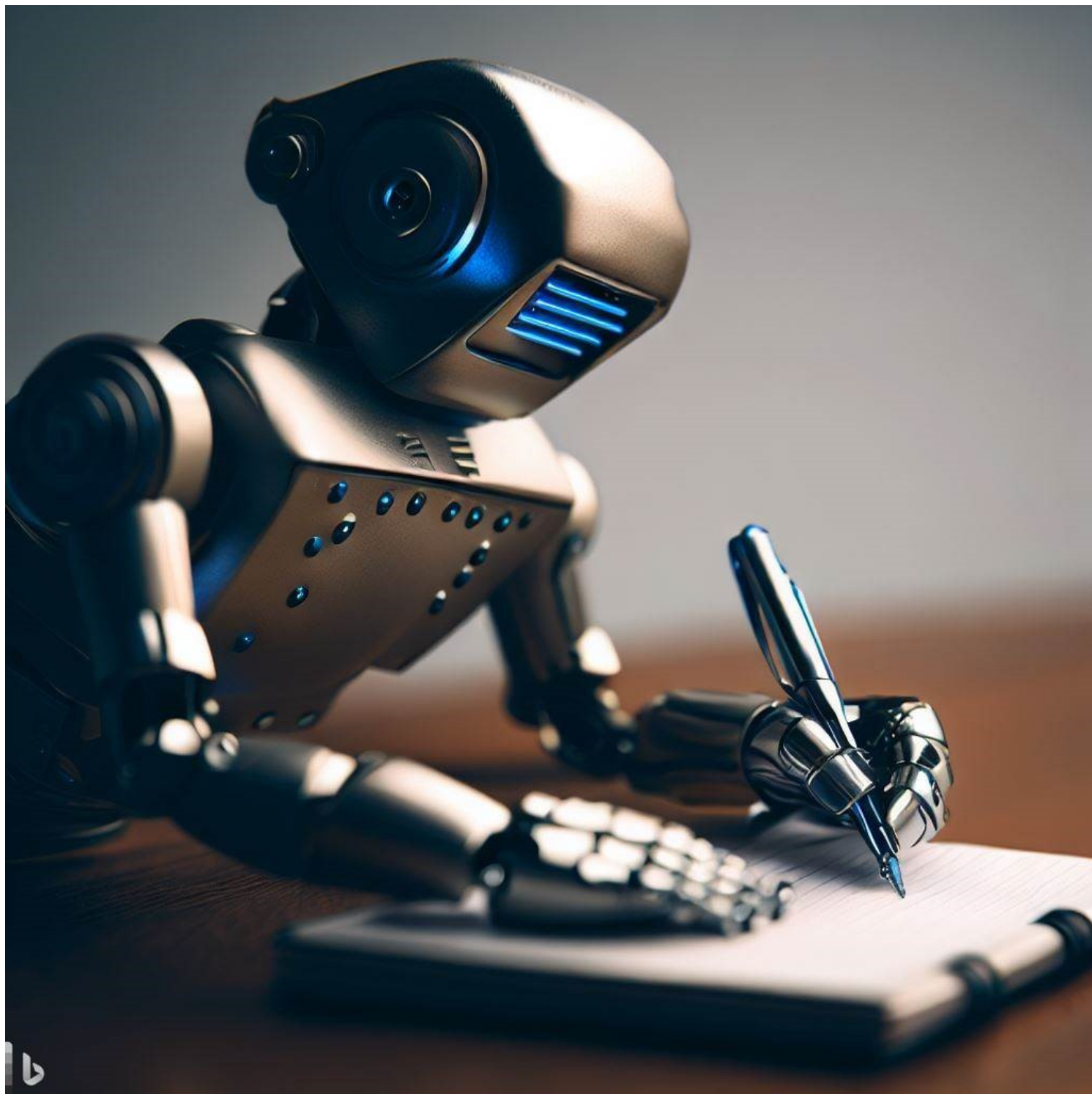


QUAND LES MÉDIAS REJOIGNENT LE PORTEFEUILLE CLIENT D'ENTREPRISES INFORMATIQUES

Publié le 6 octobre 2023



par Camille Stassart

Série : JournaBot (2/6)

Depuis le lancement, il y a près d'un an, de l'agent conversationnel ChatGPT développé par OpenAI, l'intelligence artificielle (IA) est sur toutes les lèvres. Parmi ces discours, certains expriment des craintes quant à l'avenir des métiers de la création, comme le journalisme. Ce serait toutefois oublier que de nombreuses rédactions exploitent déjà l'IA pour générer automatiquement des articles.

Depuis plusieurs années, des logiciels de production automatisée de contenus destinés aux médias sont développés par des entreprises. C'est notamment le cas des sociétés françaises Syllabs et LabSense, qui ont déjà collaboré avec plusieurs dizaines de titres de presses francophones, y compris des journaux belges.

Le NLP, un secteur de l'IA en plein essor

Comme l'indiquait l'Organisation de coopération et de développement économiques (OCDE) dans un [rapport publié en avril 2023](#), « bien que leurs déploiements n'en soient qu'à leurs débuts, les modèles de langage sont largement considérés comme étant transformateurs, ainsi qu'en témoignent la croissance rapide des investissements et l'adoption généralisée d'applications telles que l'agent conversationnel ChatGPT. »

Pour rappel, les modèles de langage font partie d'une branche de plus en plus importante de l'IA, connu sous le nom de traitement automatique du langage naturel (*Natural Language Processing* - NLP). Un domaine qui vise à développer des programmes informatiques capables de comprendre, générer et/ou traduire le langage humain.

Pour certaines entreprises proposant ce type de solution, les groupes et agences de presse représentent leurs principaux clients. Citons, à ce titre, les PME [Syllabs](#) et [LabSense](#), fondées respectivement en 2007 et 2011.

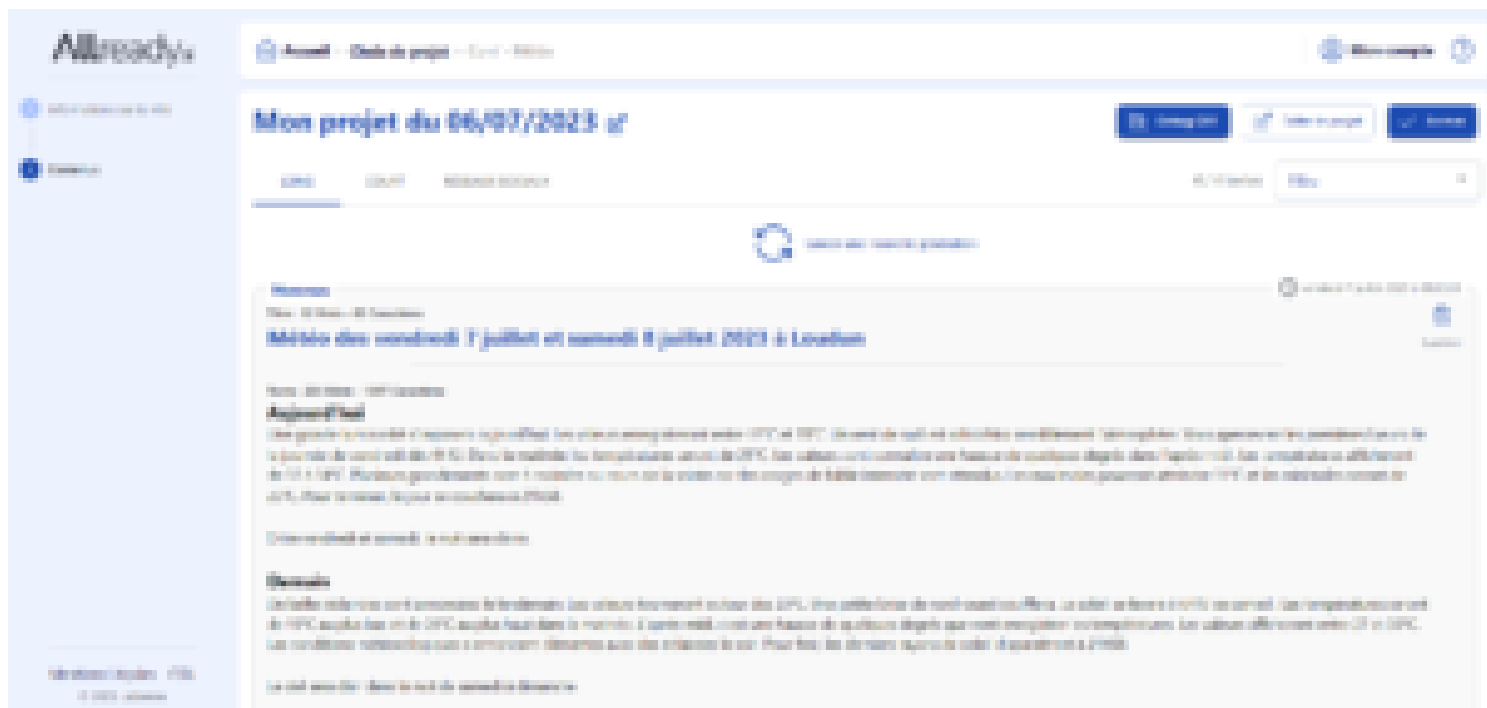
Ces sociétés ont toutes deux conçu des technologies capables de créer automatiquement des textes écrits, et se présentent comme des fournisseurs de contenus. Parmi leurs clients figurent de nombreux médias francophones comme Le Monde, Radio France, France TV, l'AFP, L'Equipe, ou encore Sudmedia, du [groupe de presse belge Rossel](#).

Des pionniers dans la génération automatisée d'articles de presse

D'après Claude de Loupy, co-fondateur de Syllabs et ancien chercheur dans le traitement des langues, si les rédactions étaient à la base frileuses vis-à-vis de ces solutions automatisées, les choses ont rapidement évolué à partir de 2015 : « En 2014, le journal [Le Monde](#) est venu nous voir pour générer des contenus automatisés lors des élections départementales de mars 2015. De cette manière, 35.000 contenus ont été produits en environ 4 heures. » Une performance qui a incité, par la suite, d'autres titres de presse à venir frapper à la porte de l'entreprise.

Aujourd'hui, les médias représentent 45% du portefeuille client de Syllabs, et environ la moitié de son chiffre d'affaires. « Pour certains, le projet est un "one shot". On a, par exemple, travaillé avec le groupe La Dépêche dans le cadre des élections municipales françaises de 2020. Mais nous n'avons (pour le moment) pas d'autre contrat avec eux. Pour d'autres, la relation s'est établie sur le long terme. On fonctionne alors par abonnement. C'est le cas pour les quotidiens Le Monde ou Ouest France. »

Au-delà des résultats d'élections, les sociétés Syllabs et LabSense sont à même de fournir des contenus automatisés en lien avec la météo, l'agenda culturel, le sport, les programmes TV, ou encore les résultats du Lotto. Des textes qui peuvent être accompagnés, si besoin, d'infographies. Pour Guillaume Desombre, CEO de LabSense – qui a jusqu'ici collaboré avec une soixantaine de médias –, la rédaction automatique de contenu vise essentiellement des « sujets pour lesquels il est économiquement difficile pour les médias d'allouer des journalistes. »



Capture d'écran montrant les prévisions météo élaborées par le logiciel LabSense © LabSense **Des logiciels réglés à la virgule près**

Pour produire ces textes, les outils développés par ces deux sociétés se fondent sur des [méthodes dites « à base de règles »](#). En clair, les logiciels reposent sur des connaissances acquises auprès d'experts humains (on parle ainsi de « systèmes-experts »), qui seront représentées par un ensemble de règles, comme des règles linguistiques, sémantiques, grammaticales, etc. Autant d'instructions qui permettront de dicter à la machine ses prises de décision. Syllabs et LabSense travaillent avec des linguistes, qui paramètrent manuellement le moteur de rédaction destiné à produire des textes.

Concernant les données sur lesquelles se basent ces contenus, les sources varient. « Dans le cas de la rédaction de contenus en lien avec des élections, on va utiliser les données fournies par le site du ministère de l'Intérieur (noms des candidats, nombre de voix, les nuls, nombre de votants...) », explique Claude de Loupy.

« Pour les contenus en lien avec le sport local, on se base plutôt sur différentes bases de données payantes fournies par des fédérations ou des sociétés spécialisées dans la collecte et l'analyse des données sportives, comme Sportradar. Il est aussi possible que le client fournisse lui-même les données dans une interface, et crée automatiquement plusieurs contenus différents », précise-t-il.

La génération automatique de textes Syllabs : comment ça marche ?



© Syllabs

La fiabilité des systèmes-experts

Cette approche à base de règles diffère des méthodes d'apprentissage automatique (*machine learning*) ou d'apprentissage profond (*deep learning*) que l'on peut retrouver derrière d'autres outils de traitement automatique du langage naturel, comme ChatGPT, qui consistent à exploiter des [algorithmes sans utiliser d'instructions explicites](#). La machine apprend donc par elle-même en étant entraînée de manière non supervisée sur d'immenses bases de données, souvent opaques.

« Avec notre approche, il n'y a justement pas ce côté « black box ». Tout est maîtrisé », indique Guillaume Desombre. « Cela permet également de ne pas avoir d'"hallucination", c'est-à-dire des réponses fausses, mais présentées comme vraies par la machine, comme cela est courant avec les modèles de *machine learning* et de *deep learning*. »

Selon les entrepreneurs, les systèmes-experts sont particulièrement fiables et les textes produits sont publiables en l'état, sans besoin d'être relus. Et si un problème ou une erreur survient avec l'outil, l'entreprise s'engage à prendre les choses en main. « C'est d'ailleurs aussi ce qui nous démarque par rapport aux grandes entreprises américaines, comme OpenIA : nous disposons d'une équipe qui accompagne et conseille les rédactions sur tous les projets », souligne le CEO de LabSense.

Des technologies hybrides à l'étude

Pour autant, les deux entrepreneurs déclarent aussi s'intéresser aux méthodes de *machine learning* et de *deep learning* : « Les systèmes-experts, basés sur des règles, restent une technologie assez lourde, dans l'idée qu'il est coûteux en temps et en argent de produire de nouveaux types de contenus », indique Claude de Loupy. De fait, les règles implémentées dans le logiciel sont spécifiques au type de contenu produit (un compte-rendu sportif, par exemple). Pour générer des contenus différents, le système doit à chaque fois être mis à jour.

« Le *deep learning*, de son côté, est une méthode plus souple. Mais elle n'est pas fiable. Aussi, il reste compliqué de produire en masse des articles en quelques heures puisque ces textes doivent être nécessairement relus. On s'applique donc aujourd'hui à coupler les deux, à utiliser l'une pour contrôler l'autre, et réciproquement. »

Même constat du côté de LabSense : « Si l'on pense toujours que les systèmes à base de règles restent les plus pertinents pour automatiser la rédaction de contenu, nous croyons aussi que coupler cette technologie au *machine learning* peut être intéressant », explique Guillaume Desombre. Cela fait ainsi quelques années que l'entreprise cherche à développer des modèles hybrides qui seront bientôt proposés aux rédactions.

Le piège de l'information artificielle

Quelle que soit la technique utilisée, les deux PDG estiment que les solutions automatisées de création de textes ont encore de beaux jours devant elles dans le secteur de l'information.

Ecoutez Guillaume Desombre en expliquer la raison :

<https://dailyscience.be/NEW/wp-content/uploads/2023/09/Son-Desombre.mp3>

Pour l'heure, rares sont encore les entreprises belges, et encore plus les entreprises de presse, à avoir intégré l'IA au cœur de leurs opérations. Et pour Nicolas van Zeebroeck, chercheur et professeur d'économie et stratégie numériques à [Solvay Brussels School of Economics and Management \(ULB\)](#), les médias ne devraient pas se précipiter vers ces nouveaux outils sans réflexion en amont. Car, selon lui, l'automatisation d'une majorité de contenus détruirait davantage

de valeur qu'elle n'en créerait.

Ecoutez Nicolas van Zeebroeck en expliquer la raison:

<https://dailyscience.be/NEW/wp-content/uploads/2023/09/Son-Van-Zeebroeck.wav>

Et le professeur de conclure : « L'erreur serait donc de voir l'IA comme un substitut ou un sous-traitant, au lieu d'un outil supplémentaire destiné à accélérer et à enrichir le travail des journalistes. Dans ce dernier cas, on pourrait arriver à quelque chose de qualitativement intéressant. »

Cette large enquête sur le journalisme automatisé par l'intelligence artificielle a bénéficié du soutien du [Fonds pour le journalisme](#) en Fédération Wallonie-Bruxelles.

JournaBot 1/6 : [Face aux IA génératives, les journalistes sont plus que jamais essentiels](#)

JournaBot 2/6 : [Quand les médias rejoignent le portefeuille client d'entreprises informatiques](#)

JournaBot 3/6 : [À l'heure de l'IA, les journalistes doivent jouer leur rôle de certificateurs de l'information](#)

JournaBot 4/6 : [Un cadre juridique pour freiner d'urgence la folle accélération de la désinformation](#)

JournaBot 5/6 : [La nature du travail journalistique en pleine mutation](#)

JournaBot 6/6 : [Une avalanche d'outils à découvrir, à maîtriser, et \(peut-être\) à utiliser](#)