



Illustration Cdi d'après photo source Vlad Alvezovsky

POINT DE VUE

# Journalisme automatisé : tour d'horizon des diverses pratiques au sein des rédactions

*Même s'il est de plus en plus question d'IA générative dans le journalisme, la plupart des rédactions utilisent une forme beaucoup plus basique de journalisme automatisé. Celle-ci a donné lieu à une grande variété d'usages. Il arrive aussi que les rédactions s'approprient la technologie à leur propre manière. Par **Samuel Danzon-Chambaud***

L'annonce par Google, l'année dernière, d'une intelligence artificielle (IA) capable de faire en partie le travail des journalistes a fait grand bruit. Selon les informations du *New York Times*, cet outil baptisé Genesis peut produire du contenu journalistique en se basant sur des détails de l'actualité. Une porte-parole de Google parle de suggestions pour des titres et des styles d'écriture, mais les cadres à qui le projet a été montré se disent déjà troublés : certains murmurent qu'il semble tenir pour acquis tous les efforts

qu'il faut pour produire de l'information fiable avec du doigté. Peu de détails ont fuité au sujet de Genesis. Selon une personne proche du dossier, Google voit dans cet outil une technologie capable d'éviter les travers des IA génératives. Il faut dire que les hallucinations des générateurs de textes et d'images peuvent laisser perplexe quant à toute utilisation d'importance dans le journalisme. En attendant de pouvoir évaluer Genesis, de nombreuses rédactions ont recours à une forme beaucoup plus basique de journalisme

automatisé, qui fonctionne à partir d'algorithmes allant chercher de la donnée (en ligne ou à l'interne) pour remplir des textes à trous rédigés à l'avance.

### *Dans certains cas, les textes ont été générés en collaboration avec le secteur universitaire.*

Dans un article scientifique<sup>1</sup> tiré de ma thèse, je recense les différentes pratiques associées à ce type de journalisme automatisé<sup>2</sup>. Je distingue essentiellement deux cas d'usage :

L'un dit *linéaire* où :

- les données proviennent de sources en position d'autorité (public ou privé) ;
- les interfaces laissent moins de place à la participation journalistique ;
- rien d'autre que du texte n'est produit.

<sup>1</sup> Samuel Danzon-Chambaud, « Automated news in practice: A cross-national exploratory study », *Open Research Europe*, vol. 3, n° 95, 2023.

<sup>2</sup> Ce projet a bénéficié d'un financement au titre du programme – cadre pour la recherche et l'innovation « Horizon 2020 » dans le cadre de la convention de subvention Marie Skłodowska-Curie n° 765140.

L'autre, où il y a *interférence* avec cette linéarité.

### **Usages linéaires**

Parmi les exemples linéaires se trouve l'agence de presse américaine Associated Press, qui collabore depuis plusieurs années avec des prestataires externes pour générer textes financiers et sportifs, à partir de données privées. De même, l'agence italienne ANSA s'en est remise à d'autres pour automatiser les chiffres du Covid-19, lesquels étaient fournis par la protection civile. France Bleu a elle aussi collaboré avec un fournisseur externe pour automatiser des résultats électoraux, utilisant pour cela des données gouvernementales.

Cette automatisation linéaire peut aussi se faire à l'interne. Des médias comme Reuters, The Times ou le diffuseur public finlandais YLE ont mobilisé leurs équipes pour produire le même type de contenu. Cette tâche incombait au personnel technique, mais aussi à des journalistes-programmeurs sachant manier pensée journalistique et pensée computationnelle. Dans certains cas, les textes ont été générés en collaboration avec le secteur universitaire. Le diffuseur public bavarois Bayerischer Rundfunk s'est ainsi associé avec l'Université technique de Munich pour automatiser la couverture de matchs de basketball.

Enfin, un dernier usage linéaire concerne la production de textes semi-automatisée. Celle-ci se fait à l'aide de logiciels externes reproduisant les principes de la programmation informatique, mais de façon suffisamment ludique pour qu'un journaliste lambda puisse les utiliser<sup>3</sup>. Plusieurs expérimentations de ce type ont été faites par la BBC, en lien par exemple avec des données sur le temps d'attente aux urgences ou sur la vitalité des grandes artères commerciales. Les possibilités d'action de l'outil sont toutefois du ressort du fournisseur.

<sup>3</sup> Ces logiciels s'apparentent à du *low code* ou du *no code*, soit des interfaces qui simplifient l'usage du code. Ce sont par exemple des outils d'édition un peu plus accessibles ou même seulement quelques interactions avec des éléments graphiques.

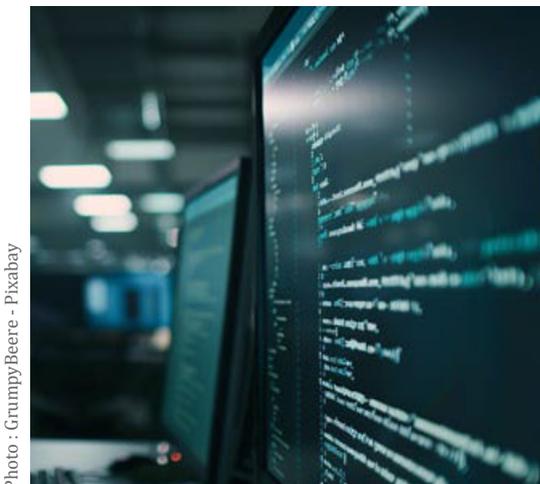


Photo : GrumpyBeere - Pixabay

## Interférences

D'autres cas d'usage interfèrent avec cette linéarité. Il s'agit là de cas à contre-courant de ceux évoqués plus haut. Ils impliquent donc l'utilisation de sources alternatives, une plus grande participation journalistique et la génération de contenus autres que du texte.

En ce qui concerne les données alternatives, la BBC et le diffuseur public australien ABC ont tous deux raccordé leur système de journalisme automatisé au fil interne qu'ils utilisent pour couvrir les résultats électoraux. Le quotidien allemand *Stuttgarter Zeitung* a lui intégré des données collaboratives sur la qualité de l'air, obtenues grâce à des capteurs citoyens.

Les réseaux sociaux ont aussi été mis à contribution, comme dans le cas d'un logiciel de Reuters qui parcourt Twitter à la recherche d'informations de dernière heure.

Une plus grande participation journalistique se fait aussi par le biais d'une production semi-automatisée, mais cette fois développée à l'interne. Reuters et le site d'informations espagnol *El Confidencial* se sont ainsi dotés de leur propre outil permettant aux journalistes de créer leurs textes de journalisme automatisé. Cela leur donne une plus grande flexibilité vis-à-vis des potentialités d'action du logiciel. Le journalisme automatisé peut également servir à produire les éléments de contexte dont les journalistes ont besoin pour leur couverture. C'est l'utilisation qu'en a faite le *Washington Post* pendant les élections américaines de 2020.

Une dernière interférence réside dans la manière dont les textes automatisés sont transformés en contenu audio, sans qu'il y ait d'intermédiaire humain. Il est question ici d'une démarche éditoriale à part entière, et non simplement d'une retranscription automatique pour personnes malvoyantes. Cela permet de créer des bulletins d'informations hyper localisés. Ils peuvent

être insérés dans des podcasts, comme dans le cas des mises à jour électorales du *Washington Post*, ou bien être diffusés via les assistants virtuels, comme ABC l'a fait avec des messages d'urgence sur Amazon Alexa.

Ces interférences témoignent d'une plus grande implication du secteur journalistique dans le domaine du journalisme automatisé. Le recours aux prestataires externes peut toutefois s'avérer judicieux lorsqu'il s'agit d'automatiser les contenus de manière ponctuelle, ou lorsque l'organisation ne dispose pas du savoir-faire technique nécessaire.

Les plateformes s'imposent aussi comme un acteur majeur, en fournissant des contenus pouvant être utilisés en

données d'entrée et en diffusant le journalisme automatisé audio sur leurs assistants virtuels. Les organisations journalistiques s'inscrivent dès lors dans une logique de dépendance, qui débouchera soit sur des négociations, soit sur un rapport de force. ■

*Samuel Danzon-Chambaud est Docteur en SIC à l'Université Dublin City.*

*Ces interférences témoignent d'une plus grande implication du secteur journalistique dans le domaine du journalisme automatisé.*