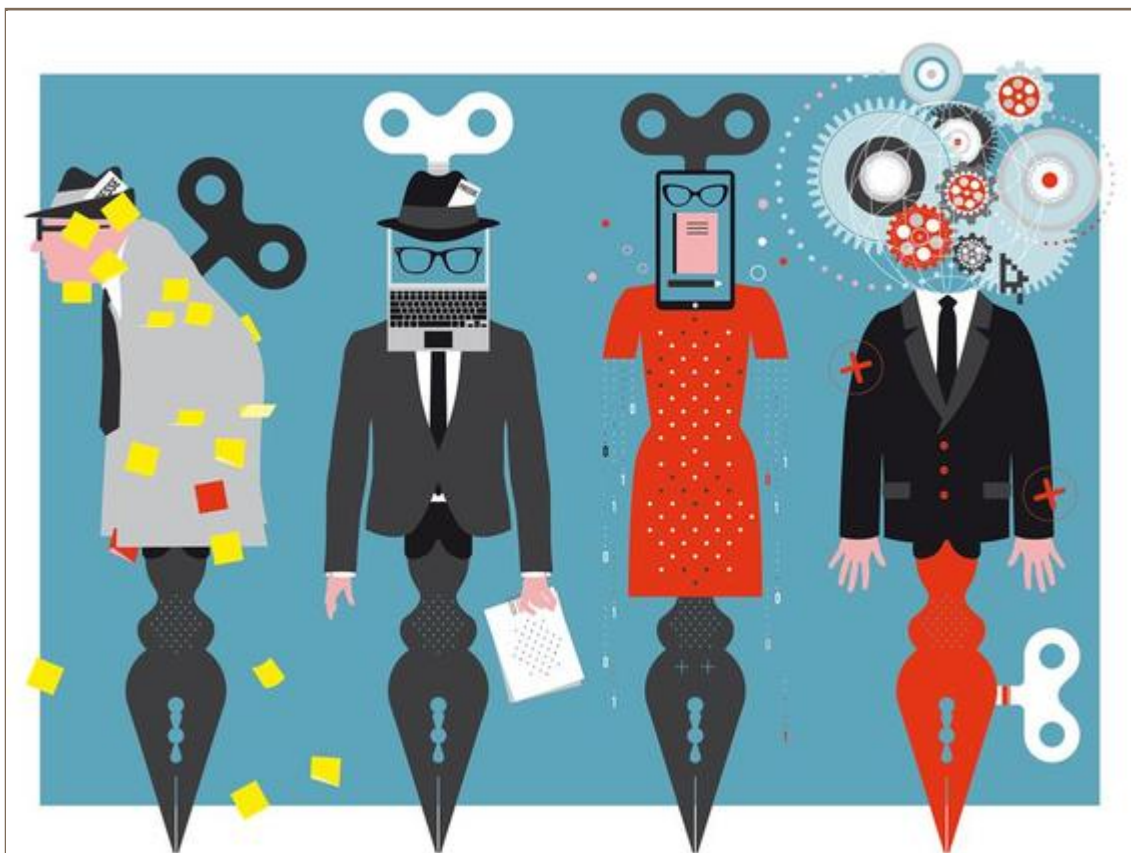


intelligence artificielle

Les algorithmes prennent la plume



Raconter un match de foot, rédiger un rapport, fournir une analyse : l'aptitude des logiciels à écrire n'est plus de la science-fiction. Les professions intellectuelles vont-elles devenir leurs supplétifs ?



Le 17 mars à l'aube, les habitants de Los Angeles sont réveillés par un léger tremblement de terre. Moins de trois minutes plus tard, le site Internet du quotidien *Los Angeles Times* publie un premier article sur le sujet - à première vue, une dépêche rédigée dans l'urgence par une agence de presse : " *Un séisme peu profond de magnitude 4,7 a été signalé lundi matin à cinq miles de Westwood, en Californie, selon l'US Geological Survey. La secousse s'est produite à 6 h 25, heure du Pacifique, à 5 miles sous la surface terrestre. (...) L'épicentre était à 6 miles de Beverly Hills, à 7 miles de Universal City (...). Au cours des dix jours précédents, il n'y avait pas eu de séisme de magnitude 3 ou supérieure dans cette zone.* " En revanche, la signature est inhabituelle : " *Cette information provient du service de notification de séismes de l'US Geological Survey, et l'article a été créé par un algorithme écrit par l'auteur.* "

L'auteur en question est un employé du *Los Angeles Times*, à la fois journaliste et programmeur informatique. Ce matin-là, les serveurs de l'US Geological Survey (USGS) ont reçu les informations provenant des sismographes et les ont transformées en données chiffrées, qu'ils ont ensuite transmises par Internet à l'ordinateur personnel du journaliste. Là, les données ont été aspirées par un petit programme qui a sélectionné les informations pertinentes, puis rédigé un article en anglais courant. Ce jour-là, le journaliste, réveillé en sursaut, s'est levé et a lu l'article avant de cliquer sur " Envoi ", mais en fait ce genre de texte pourrait être publié sans intervention humaine. La prochaine fois, le journaliste pourra rester couché, ses lecteurs seront quand même informés.

Cet événement a attiré l'attention des médias américains, car ces dernières années le *Los Angeles Times* a licencié de nombreux journalistes, à cause de problèmes financiers. De là à imaginer qu'ils ont été remplacés par des machines, il n'y a qu'un pas... En fait, des " robots écrivains ", plus ou moins élaborés et autonomes, commencent à être installés, en toute discrétion, dans quelques médias et dans divers secteurs d'activité gros producteurs de documents écrits.

Aux Etats-Unis, la technique de l'écriture automatisée a été en partie développée par les chercheurs en intelligence artificielle de l'université Northwestern, près de Chicago. Le professeur Larry Birnbaum, codirecteur du Laboratoire d'information intelligente, incarne à lui seul cette nouvelle discipline transversale, car il enseigne aussi à l'école de journalisme Medill, voisine de Northwestern.

Il est l'un des principaux inventeurs du système Quill (" plume ") qui, malgré sa puissance, tient dans un ordinateur Linux de modèle courant. Pour expliquer son fonctionnement, Larry Birnbaum distingue quatre étapes théoriques, même si, en pratique, elles se chevauchent.

Tout d'abord, Quill reçoit d'une base de données externe des informations structurées par un logiciel (tableaux, listes, graphiques). *" De nos jours, cela représente une grande partie de l'information produite par les humains, depuis le tableur contenant la comptabilité d'une entreprise jusqu'au blog résumant le déroulement d'un match de football. "* En amont, d'autres systèmes intelligents peuvent se charger de transformer des informations sous des formats disparates (y compris des textes) en données structurées exploitables par une machine. Ainsi, les robots écrivains ont potentiellement accès à l'ensemble du savoir humain.

La deuxième tâche, accomplie par des algorithmes, consiste à effectuer une " analyse narrative " : *" Les données sont triées et ordonnées selon une méthode strictement orientée vers la construction d'une narration, explique Larry Birnbaum. Elle sélectionne certains faits, souligne certaines actions, met en valeur certains chiffres... "*

La troisième, la plus innovante, est la " genèse de la narration " : *" Les algorithmes établissent un plan, une liste de faits, poursuit Larry Birnbaum. Puis, grâce à la modélisation, ils choisissent des angles éditoriaux. En pratique, le résultat est un mélange de mots, de lignes de code, de graphiques - une représentation compréhensible uniquement par les machines. "*

A partir de ces éléments, Quill peut passer à l'écriture proprement dite : *" Pour composer ses phrases, il possède une bibliothèque de règles, de mots et d'expressions - de l'anglais courant, et aussi du vocabulaire professionnel spécialisé ",* précise Larry Birnbaum. Pour les non-initiés, cette ultime étape est la plus spectaculaire et la plus étonnante, mais pour les professionnels de l'intelligence artificielle, ce n'est pas la plus complexe. *" Les ordinateurs savent écrire en anglais depuis longtemps. S'ils ne le faisaient pas, c'est qu'ils n'avaient rien à nous dire, puisqu'ils n'avaient pas accès à une masse suffisante d'informations ",* s'amuse le professeur. Désormais, ils ont enfin des histoires intéressantes à raconter aux humains.

Pour exploiter Quill commercialement, un autre chercheur du Laboratoire d'information intelligente de Northwestern, Kris Hammond, a créé une start-up baptisée Narrative Science, installée dans le centre de Chicago. Kris Hammond est persuadé que, contrairement aux idées reçues, un bon discours vaut mieux qu'un beau croquis, car le cerveau humain est câblé pour comprendre parfaitement une idée quand on la lui explique avec des mots. " Grâce à Quill, dit-il dans quelques années, plus personne ne sera obligé de s'escrimer à déchiffrer un tableur Excel ni à interpréter un graphique avec des abscisses et des ordonnées... Quill et ses successeurs absorberont ces données indigestes et les transformeront en textes clairs et simples qui permettront à tout le monde de tout comprendre de la façon la plus naturelle qui soit : par le langage. "

Récemment, Kris Hammond s'est fait remarquer en affirmant que, d'ici à 2025, 90 % des informations lues par le grand public seraient générées par des ordinateurs : " Cela ne signifie pas que 90 % des journalistes seront remplacés par des robots, mais que le volume de textes publiés va croître de façon gigantesque, explique-t-il. Prenons l'exemple des petits matchs de base-ball amateurs : ils n'intéressent pas les médias, mais chacun d'eux est suivi avec passion par quelques dizaines de personnes. Quill se

procure les données de jeu sur des milliers de ces matchs et produit en un instant des milliers d'articles, un par match, dans un style semblable à celui des journalistes sportifs, qui est facile à imiter. " Narrative Science possède déjà des clients dans ce secteur : des sites spécialisés dans le sport local ou dans les informations pour la jeunesse.

Quill s'intéresse aussi aux finances, autre secteur où les articles écrits par des humains sont souvent répétitifs. " *Depuis longtemps, le magazine Forbes publie des prévisions de bénéfices pour certaines entreprises, avant la publication de leurs résultats. Désormais, grâce à Quill, il le fait pour plus de 5 000 sociétés* ", révèle Kris Hammond. On peut désormais lire sur Forbes.com des textes signés Narrative Science : " *Wall Street est optimiste sur les perspectives de la société Kruger, qui devrait publier ses résultats pour le premier trimestre dès jeudi prochain. Les analystes prévoient un bénéfice de 1,05 dollar par action, soit une augmentation de 92 % par rapport à l'année dernière...* " De même, des banques, des courtiers et des agences de notation, qui préfèrent rester anonymes, utilisent Quill pour rédiger les innombrables rapports exigés par l'administration et diverses agences de contrôle : " Aujourd'hui, les rapports écrits par Quill sont relus avant d'être envoyés, car nous sommes encore en phase expérimentale. Mais, dans quelques mois, ils seront expédiés automatiquement à l'administration, sans avoir été vus par un œil humain ", explique-t-il.

Selon Kris Hammond, Quill est le point de départ d'une grande aventure. L'écriture automatisée va bientôt donner toute sa mesure, en se mariant à une autre avancée technologique : le traçage individuel de milliards de consommateurs, grâce à leurs achats, leur navigation Internet, leurs communications... : " *Un jour, le président Obama a expliqué dans les médias que, si les Américains gonflaient mieux les pneus de leur voiture, ils réduiraient leurs dépenses de carburant de 7 %. Cet argument n'a pas porté, car les gens n'avaient pas envie de faire le calcul. Dans l'avenir, votre site d'information favori vous expliquera les choses différemment : il saura qui vous êtes, il connaîtra la marque de votre voiture et sa consommation, la distance que vous parcourez chaque semaine, le type de carburant que vous achetez, etc. Il rédigera à votre intention un article personnalisé qui indiquera exactement combien de dollars vous économiseriez si vos pneus étaient mieux gonflés.* " On peut multiplier les exemples dans tous les secteurs, de la santé à la politique : " *A l'avenir, chaque article aura un seul lecteur.* "

Narrative Science n'est pas seul sur le marché du robot écrivain : les initiatives concurrentes se multiplient. Ainsi, la société Automated Insights, installée en Caroline du Nord, vend un système baptisé Wordsmith (" l'artisan de mots "), décrit comme une " *plate-forme de génération de langage naturel* ". Son directeur commercial, Adam Smith, affirme que sa société a produit plus de 300 millions de textes en 2013 et va dépasser le milliard en 2014. Il compte une douzaine de clients en phase expérimentale, dont le groupe de presse Gannett, qui publie *USA Today* et Yahoo News.

Par ailleurs, Yahoo ! utilise Wordsmith pour rédiger à la chaîne les textes de son jeu Fantasy Sport - le joueur crée l'équipe de football de ses rêves en utilisant les profils professionnels de vrais sportifs et affronte les équipes virtuelles des autres joueurs dans des matchs fictifs. " *Le système analyse les performances des sportifs dans les vrais matchs, puis il décide quelle équipe virtuelle va gagner contre telle autre. Et, bien sûr, il raconte le match* ", précise Adam Smith.

Wordsmith sait aussi rédiger des annonces immobilières, des rapports marketing, des bilans d'activité ou des comptes rendus de performances financières. " *Avec les mêmes données - les cours de la Bourse -, nous pouvons écrire des millions d'histoires différentes, chacune centrée sur l'évolution d'un portefeuille individuel* ", note Adam Smith.

De son côté, la société française Yseop, installée en partie au Texas, a mis au point un système capable de parler anglais, français, espagnol, portugais et bientôt japonais. A titre de test, Yseop propose sur son site Internet un article financier (en anglais) dont les phrases s'adaptent automatiquement à chaque fois qu'on modifie un chiffre dans un paragraphe précédent. Au départ, l'article note " *une évolution franchement positive des dépôts bancaires aux Etats-Unis* ", mais quand on inscrit un chiffre plus bas le texte se corrige pour parler, selon le cas, de " *baisse marginale* " ou de " *déclin substantiel* ".

Selon son patron, Jean Rauscher, Yseop travaille pour les services clients de banques et de compagnies de télécoms et pour des sites d'information appartenant à des sociétés financières. Il fournit aussi un

service pointu à des agences d'investigation spécialisées dans les affaires boursières. " *Pour en savoir plus sur un dirigeant d'entreprise, notre système va scanner une trentaine de bases de données et recueillir des informations à son sujet, y compris des éléments de son casier judiciaire, explique Jean Rauscher. Puis son CV est rédigé automatiquement.* "

Les robots écrivains pourraient se banaliser rapidement, car Yseop permet à ses clients d'acheter son système et de l'exploiter directement en interne. A terme, toutes les grandes entreprises, qui sont obligées de produire des séries de rapports de plus en plus volumineux sur tous les aspects de leurs activités, sont des clients potentiels. Jean Rauscher essaie déjà d'imaginer une collaboration harmonieuse entre le robot et l'employé de bureau : " *Yseop prépare un plan, ou un brouillon, contenant les informations essentielles, avec les chiffres exacts. Ensuite, un humain prend le relais, il retravaille le brouillon, en rajoutant un peu d'opinion et de jugements de valeur.* "

Le système sait aussi dialoguer avec les humains, qui deviennent de facto ses assistants : " *Si l'algorithme s'aperçoit qu'il lui manque une donnée, il s'arrête et demande qu'on la lui fournisse. Dès qu'il l'obtient, il se remet au travail.* " Chaque employé de bureau pourra alors se demander si ses compétences sont complémentaires de celles de son ordinateur, ou si l'ordinateur se débrouillerait mieux sans lui.

Yves Eudes

© Le Monde

◀ **article précédent**
Biobanques

article suivant ▶
Marketing, la joute des programmes