

Thélyson Orélien : c'était l'IA ou mourir ?

Favori dans la course aux prix, « C'était ça ou mourir » du romancier d'origine haïtienne Thélyson Orélien est suspecté d'avoir été rédigé par intelligence artificielle. Trois questions pour comprendre.

DÉCODAGE
CÉDRIC PETIT

La rentrée littéraire avait déjà son phénomène et son favori pour la saison des prix (Goncourt en tête). Elle a désormais aussi sa polémique, cristallisée autour du même roman et du même auteur : Thélyson Orélien et son livre, paru chez Grasset fin août, *C'était ça ou mourir*. Repris dans les premières sélections des principaux prix littéraires, trônant au sommet des ventes de roman, celui-ci a été récompensé, lundi, du prix Fnac du roman. Décerné par 400 adhérents et autant de libraires de l'enseigne, il a pour vocation de « célébrer la diversité de la création littéraire contemporaine, qu'elle soit francophone ou traduite ». Un triomphe de plus pour cet ouvrage paru en mars aux éditions du Boréal, à Québec, avant de traverser les frontières en ce mois de septembre, non sans s'être déjà fait remarquer au Canada, et valoir à son auteur, Thélyson Orélien, un concert de louanges et un portrait dans le *New York Times*.

1 D'où vient le doute ?
Un triomphe qui a partiellement été gâché ou terni par des accusations qui ont fleuri le jour même où le prix Fnac lui était décerné. En première ligne, un compte X, Balance Ton Claude, du nom du modèle de langage du géant Anthropic, qui a lancé la charge : soumis à un détecteur d'IA, le texte du roman ne serait pas l'œuvre de l'écrivain. « Ce roman est rédigé presque intégralement par une intelligence artificielle. Nous avons testé plusieurs passages, à divers endroits du manuscrit, pour presque toujours le même résultat : 100 % IA », affirme le responsable du compte X, saisies d'écran à l'appui. Le constat s'appuie sur l'analyse et les chiffres livrés par l'outil Pangram, qui se revendique être le plus performant dans le domaine de la détection de contenus IA. « N'aurait-il pas été plus honnête que l'auteur crédite son co-auteur ? », interroge ainsi le compte anonyme, en conclusion du post.

Vendu à 55.000 exemplaires en trois semaines depuis sa sortie, *C'était ça ou mourir* n'aurait donc pas pour seul auteur Orélien Thélyson, d'origine haïtienne comme son personnage principal Jonas Dorléon. Le constat, mécanique, effectué par Pangram, ne pouvait qu'éveiller la fachosphère. A l'image du journaliste Paul Sugy, étiqueté de droite, le premier à dénoncer le « pot

aux roses » et à défendre le travail du détecteur de texte Pangram « relativement fiable » : « Quand il suspecte qu'un texte a été généré par IA, il se trompe rarement », écrivait-il, rapidement relayé à son tour par les réseaux d'extrême droite. La critique littéraire française et le monde de l'édition n'y auraient vu, se réjouit-on ainsi sur les réseaux sociaux, que du feu, inondant de louanges et sélectionnant le roman sur les listes du Goncourt, Renaudot, Femina, Médicis, Goncourt des lycéens.

2 L'outil de détection de Pangram est-il un outil fiable ?
D'Olivier Nora, ancien directeur de Grasset, aux libraires français qui lui ont décerné, lundi également, le prix *Livres Hebdo*, tous dupés ? Pour Katrien Beuls, chercheuse à l'Université de Namur, la fiabilité de l'outil Pangram ne laisse planer qu'un doute assez mince. Avec un taux d'IA de 100 %, précise-t-elle, il y a très peu de possibilités que la prédiction donnée par Pangram soit elle-même fausse. « En 2023 ou en 2024, on aurait pu prendre ces résultats moins au sérieux. Aujourd'hui, non. » Mais les exemples de Victor Hugo et de la Bible, que Pangram aurait à tort détectés comme produits par l'IA ? « C'est précisément le cœur de l'outil : il ne pouvait pas déterminer si la Bible était écrite par un être humain, parce qu'il n'existe pas d'autres textes de cette période. Comment fonctionne Pangram ? Le texte qu'on lui soumet est comme "jeté" sur un corpus très large de textes et découpé en "fenêtres". Pour chacune, il étudie si celles-ci ressemblent à des textes, qu'ils soient générés par des machines ou des êtres humains. Dans cet ensemble, il recherche des caractéristiques existantes, ce qui lui permet de déduire si le texte est ou non un *outlier*, soit une valeur extrême. »

3 Comment l'auteur se défend-il ?
Contactées par Radio France, les éditions Grasset ont d'abord refusé de réagir « face à une campagne de haine », avant de lancer la contre-offensive : « Ces outils censés reconnaître l'IA sont eux-mêmes des IA qui fabulent que la Bible a été écrite artificiellement. Comment faire confiance à ces outils ? », interrogent-elles, en cours de journée. L'auteur lui-même, Thélyson Orélien, qui dit être victime de nombreuses attaques racistes, a aussi pris la parole, via son éditeur. « J'ai commencé à écrire



C'était ça ou mourir à l'automne 2017 et j'en ai achevé le premier jet en 2019. Je viens d'une tradition littéraire haïtienne et caribéenne où la poésie, les images, les répétitions et le rythme portent une parole vivante. C'est dans cet héritage que ma voix s'est construite. On peut discuter de mon style, mais faire de ses procédés des preuves d'intelligence artificielle n'a rien à voir avec la critique littéraire. Je me demande si l'on ne soupçonnerait pas aussi Césaire ou Jacques Stephen Alexis, dont je tente modestement d'appréhender la liberté de langue. »

En réponse aux critiques émises sur son constat initial, le responsable anonyme du compte Balance ton Claude a de nouveau fait tourner le logiciel Pangram, à qui il a soumis d'autres textes de Thélyson Orélien, parus avant *C'était ça ou mourir*. Incomplets, ceux-ci afficheraient un taux d'IA de 0 %... « La probabilité que des dizaines d'extraits de son roman soient identifiés à tort comme étant générés à 100 % par une IA est presque nulle (elle est d'un sur plusieurs trilliards de trilliards de trilliards), statistiquement indistinguable de zéro. » Ce qui tendrait à démontrer que les textes rédigés par le Québécois avant l'arrivée des IA étaient

« On peut discuter de mon style, mais faire de ses procédés des preuves d'intelligence artificielle n'a rien à voir avec la critique littéraire », se défend Thélyson Orélien. © AFP.

bien de sa plume, contrairement à *C'était ça ou mourir*. Pour Katrien Beuls encore, ces autres tests menés « ne plaident pas en faveur de Thélyson Orélien. Si tous ses textes étaient détectés comme IA, on pourrait se dire que Pangram échoue à reconnaître un style, une plume avant-gardiste. Ce n'est pas le cas pour le livre, à moins qu'il n'ait totalement changé de style », relève-t-elle. « Des livres écrits par des modèles de langage, il en existe aujourd'hui des milliers en vente sur Amazon. Le grand public doute encore de la fiabilité des outils de détection. Mais aujourd'hui, ils sont très au point. » Après avoir elle-même passé l'intégralité du texte à la loupe de Pangram – Balance ton Claude n'avait fait passer que des extraits à la moulinette –, elle observe un score de 92 %. Avec des passages « jugés humains ». Quand le texte comporte deux espaces entre « 8 » et « h », la chercheuse y voit la marque d'un travail humain : « Dans le reste du texte, il n'y en a qu'un. Une machine n'intégrerait jamais deux espaces. »

La maison d'édition du Boréal, à Québec, a déclaré à l'AFP avoir « pris connaissance des allégations » et ne pas les « prendre à la légère », tout en maintenant la confiance « envers le travail de Thélyson Orélien (...) entière »...

En 2023 ou en 2024, on aurait pu prendre ces résultats moins au sérieux. Aujourd'hui, non

Katrien Beuls
Chercheuse à l'Université de Namur



20018630

Achète à très bon prix

Achat Fourrures

Manteaux de fourrure: vison, astrakan, renard,...

Argentierie: couverts et pièces de forme | Armes anciennes: fusil, pistolet, épée, sabre | Montre gousset / bracelet | Instruments de musique: piano, violon, saxo,... | Livres anciens: dictionnaire, BD, missel,... | machine à coudre et poste radio | Meubles et objets anciens: pendule, tableau, sculpture, miroir, luminaire,... | Bijoux or, argent, fantaisie,... | Pièces de monnaies anciennes / Cuivre et étain



CHARLES Anthony | 0484/20 26 78 |

6