Sur les limites de l'identification par l'humain de textes générés automatiquement

Nadège Alavoine, Maximin Coavoux, Emmanuelle Esperança-Rodier, Romane Gallienne, Carlos-Emiliano González-Gallardo, Jérôme Goulian, José G. Moreno, Aurélie Névéol, Didier Schwab, Vincent Segonne, et al.

Sur les limites de l'identification par l'humain de textes générés automatiquement

Nadége Alavoine[1], Maximin Coavoux[2], Emmanuelle Esperança-Rodier[2], Romane Gallienne[3], Carlos-Emiliano González-Gallardo[4], Jérôme Goulian[2], Jose G. Moreno[5], Aurélie Névéol[6], Didier Schwab[2], Vincent Segonne[7] and Johanna Simoens[8]

(1) Université Paris-Saclay, LISN, Campus Universitaire bâtiment 507, Rue du Belvédère, 91400 Orsay, France
(2) Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
(3) Université Sorbonne Nouvelle, Lattice, CNRS, ENS-PSL, 1 rue Maurice Arnoux, 92120 Montrouge, France
(4) La Rochelle Université, L3i, 17000 La Rochelle, France
(5) University of Toulouse, IRIT, 31000 Toulouse, France
(6) LISN, Université Paris-Saclay, CNRS, 91403 Orsay, France
(7) Université Bretagne Sud, UMR CNRS 6074, IRISA, F-56000 Vannes, France
(8) Everteam, Bagneux, France

nadege.alavoine@universite-paris-saclay.fr,
{first.last}@univ-grenoble-alpes.fr
romane.gallienne@cnrs.fr, carlos.gonzalez_gallardo@univ-lr.fr,
jose.moreno@irit.fr, aurelie.neveol@lisn.upsaclay.fr
vincent.segonne@univ-ubs.fr, johanna.simoens@gmail.com

RÉSUMÉ

La génération de textes neuronaux fait l'objet d'une grande attention avec la publication de nouveaux outils tels que ChatGPT. La principale raison en est que la qualité du texte généré automatiquement peut être attribuée à un·e rédacteurice humain·e même quand l'évaluation est faite par un humain. Dans cet article, nous proposons un nouveau corpus en français et en anglais pour la tâche d'identification de textes générés automatiquement et nous menons une étude sur la façon dont les humains perçoivent ce texte. Nos résultats montrent, comme les travaux antérieurs à l'ère de ChatGPT, que les textes générés par des outils tels que ChatGPT partagent certaines caractéristiques communes mais qu'ils ne sont pas clairement identifiables, ce qui génère des perceptions différentes de ces textes par l'humain. Ceci est le résumé de l'article "Limitations of Human Identification of Automatically Generated Text" publié à LREC-COLING-2024 (Alavoine *et al.*, 2024).

ABSTRACT

Here the title in English.

Neural text generation is receiving broad attention with the publication of new tools such as ChatGPT. The main reason for that is that the achieved quality of the generated text may be attributed to a human writer by the naked eye of a human evaluator. In this paper, we propose a new corpus in French and English for the task of recognising automatically generated texts and we conduct a study of how humans perceive the text. Our results show, as previous work before the ChatGPT era, that the generated texts by tools such as ChatGPT share some common characteristics but they are not clearly identifiable which generates different perceptions of these texts.

MOTS-CLÉS : identification humaine, génération de texte avec des modèles neuronaux, ChatGPT.

KEYWORDS: human identifying, neural text generation, ChatGPT.

Références

ALAVOINE N., COAVOUX M., ESPERANÇA-RODIER E., GALLIENNE R., GALLARDO C. G., GOULIAN J., MORENO J. G., NEVEOL A., SCHWAB D., SEGONNE V. *et al.* (2024). Limitations of human identification of automatically generated text. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, p. 10511–10516.