ChatGPT and Spontaneous Metacognition: Analyzing AI's Classification of Subjective Cognitive Experiences

Gull Zareen[1], Coline Grégoire[2], Chris J.A. Moulin[1], Akira R. O'Connor[3]
[1]Laboratoire de Psychologie & NeuroCognition (LPNC CNRS 5105), Université Grenoble Alpes, France
[2]Centre de Recherche sur la Cognition et l'Apprentissage (CeRCA CNRS 7295), Université de Tours, France
[3]School of Psychology and Neuroscience, University of St. Andrews, Scotland

Abstract

Involuntary cognition includes subjective phenomena like déjà vu, tip-of-the-tongue (TOT) state and jamais vu. Given their spontaneity and subjectivity, it remains challenging to understand how these phenomena are perceived and classified. In this paper, we take an existing dataset from a reasonably large sample (n=860) across three countries (Poland, France and Pakistan) and evaluate ChatGPT's ability to classify subjective descriptions of spontaneous experiences. We compared ChatGPT's responses (versions 3.5 and 4.0) with human participants to assess its accuracy in classification. While ChatGPT showed high accuracy for déjà vu and TOTs, it struggled with jamais vu, indicating limitations in classifying less well-defined spontaneous cognition.

Keywords : spontaneous, metacognition, déjà vu, ChatGPT

1 Introduction

Spontaneous metacognition can be defined as thoughts which automatically come to mind which signal feelings and processes relevant to cognitive processing (e.g. déjà vu, jamais vu, tip-of-the-tongue (TOT). Due to their relative infrequency, spontaneity and subjectivity, most cognitive psychology research relies on questionnaires asking people to retrospectively evaluate their past experiences (e.g. O'Connor et al., 2021; Metcalfe et al., 2017). This study focuses on one such retrospective questionnaire dataset, and examines the extent at which we can use AI (ChatGPT) to better understand the classification of six spontaneous subjective experiences.

We were interested in déjà vu (the feeling of familiarity for something known to be a unique, first-time experience) jamais vu (the feeling of unreality or strangeness for a stimuli known to be familiar), TOT (a sensation of the inability to retrieve information), error detection (spontaneously realizing one has made an error), involuntary autobiographical memory (IAM; a personal past experience coming to mind spontaneously) and zoning out (realizing that one's attention has wondered from the central goal).

ChatGPT's abilities extend from addressing simple inquiries to facilitating challenging discussions about productivity in nearly human-like language (Liu et al., 2021). In the current study, we wanted to investigate if we can rely on AI tool like ChatGPT to analyse subjective human experiences (AI as qualitative analysis tool see Sen et al., 2023; Morgan, 2023). Our reasoning was that ChatGPT's extensive training renders it useful to researchers for classifying and verifying subjective reports based on a synthesis of extant knowledge and language use. Hence, in this study we analyse the accuracy rate of ChatGPT in classifying the six spontaneous experiences above.

2 Methodology

The subjective descriptions of six phenomena used in this study were collected from a large online survey (n=860) across three countries (Poland, France and Pakistan - see the pre-registration: https://osf.io/h8k9z/). We tested both ChatGPT versions 3.5 and 4.0 with these descriptions in the current study. Our study has three experiments. In

Experiment 1, we fed ChatGPT (4.0) subjective descriptions of déjà vu and jamais vu in French and English language and analysed how accurately ChatGPT could classify these experiences. In Experiment 2, we fed ChatGPT (3.5) subjective French descriptions from six different spontaneous phenomena and some control descriptions. We first fed these descriptions in a blocked order and then in a randomized order and analysed the classification accuracy. In Experiment 3, French human participants were also asked to classify these experiences based on the same set of descriptions and then we cross-validated findings of Experiment 2 with those of human participants.

3 Results

ChatGPT excelled in classifying TOT (100% accuracy) and déjà vu (85% accuracy) descriptions. However, ChatGPT performed poorly in categorizing jamais vu (5% accuracy) and our control descriptions (15% accuracy). The human participants' accuracy in classifying jamais vu (d'= 2.87) and control descriptions (d'=2.20) was higher than ChatGPT as shown in Table 1, suggesting that as yet human cognitive experiences are fundamentally difficult for AI models to grasp fully.

Experience	Exp. 2 ChatGPT 3.5 (d', c)	Exp. 3 Humans (d', c)	Exp. 3 ChatGPT 4.0 (d', c)
Déjà vu	2.30, 0.11	2.65, 0.74	2.53, 0.66
TOT	4.35, -0.91	3.99, 0.76	3.50, -0.28
Jamais vu	0.47, 1.88	2.87, 0.53	0.46, 1.23
IAMs	2.49, 0.48	2.73, 0.54	3.06, 0.43
Error detection	1.89, 0.27	2.42, 1.01	2.13, 0.46
Zoning out	2.33, 0.40	2.92, 0.32	3.01, 0.60
Control statements	1.35, 1.71	2.20, 0.08	1.92, 0.54

Table 1: Sensitivity (d') and bias (c) in classifying spontaneous metacognitive experiences

4 Discussion

The overall performance of ChatGPT (71% accuracy) in analysing and categorizing subjective descriptions of six phenomena was similar to human participants (79% accuracy). This suggests that AI tools like ChatGPT may be efficient in analysing intensely subjective experiences such as déjà vu and other spontaneous metacognitive experiences presented here.

However, ChatGPT constantly struggled in recognizing/classifying jamais vu and control descriptions. This problem could be due to a lack of sufficient training data (and even the lack of a coherent definition in the literature), and subjective and complex nature of these experiences. This demonstrates that AI currently lacks a complete understanding of more nuanced cognitive phenomena.

5 Conclusion

ChatGPT here offered a useful means of summarizing a large qualitative dataset, an exciting possibility for cognitive psychologists. To rely more fully on AI models like ChatGPT as an analytical tool for analysing large qualitative data, particularly classifying subjective (meta)cognitive experiences, requires further development and improvement in the training dataset, something which is shown in its classification of jamais vu experiences.

References

O'Connor, A. R., Wells, C., & Moulin, C. J. A. (2021). Déjà vu and other dissociative states in memory. *Memory*, *29*(7), 835–842. https://doi.org/10.1080/09658211.2021.1911197

Metcalfe, J., Schwartz, B. L., & Bloom, P. A. (2017). The tip-of-the-tongue state and curiosity. *Cognitive research: principles and implications*, *2*, 1-8.

Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z., & Tang, J. (2021). GPT understands, too. *AI Open*. https://doi.org/10.48550/arXiv.2103.10385

Sen, M., Sen, S.N., & Sahin, TG (2023). A New Era for Data Analysis in Qualitative Research: ChatGPT!. *Shanlax International Journal of Education*, *11*, 1-15.

Morgan, D. L. (2023). Exploring the use of artificial intelligence for qualitative data analysis: The case of ChatGPT. *International journal of qualitative methods*, *22*, 16094069231211248.