Marmara Üniversitesi Öneri Dergisi • Cilt 19, Sayı 62, Temmuz 2024, ISSN 2147-5377, ss. 1-17
DOI: 10.14783/maruoneri.1363643

ARAŞTIRMA MAKALESİ / RESEARCH ARTICLE

DRM LISTS CREATED WITH CHATGPT: ANALYSING RECOGNITION MEMORY WITH CLIMATE-CHANGE THEMED LISTS* **

CHATGPT İLE OLUŞTURULAN DRM LİSTELERİ: İKLİM DEĞİŞİKLİĞİ TEMALI LİSTELERLE TANIMA BELLEĞİ ANALİZİ

Gökhan ŞAHİN***

Abstract

Machine learning and natural language processing have led to the development of powerful language models such as ChatGPT, which can generate consistent and human-like responses to a wide range of queries. In many domains, ChatGPT provides appropriate responses to given commands. One of the aims of this study is to investigate the use of these association lists, such as the Deese-Roediger-McDermott (DRM) lists popular in cognitive psychology studies, by ChatGPT by giving the necessary instructions. The same method was then used to create association lists around a specific topic (climate change). The results of the first study showed that participants gave more false answers when discriminating whether critical words were presented during the test phase than when related and unrelated words were presented. This finding shows that DRM lists generated by ChatGPT can be used to search for memory errors. In line with the literature, false answers for critical words were predominantly rated as 'remember'. The results of the second study, which was applied to the lists created

* This study is an expanded version of the paper presented by the author at the 4th International Black Sea Modern Sıientific Research Congress.

** The survey in this article was approved by the decision number 2023/234 taken at the meeting of Recep Tayyip Erdoğan University Social Sciences and Humanities Ethics Committee on 25/08/2023 and numbered

*** Karadeniz Technical University, Department of Psychology, gsahin@ktu.edu.tr ORCID 0000-0002-6479-7018.

How to cite this article/Atıf için: Şahin, G. (2024). DRM lists created with chatgpt: Analysing recognition memory with climate-change themed list. *Öneri Dergisi, 19*(62), 1-17. DOI: 10.14783/maruoneri.1363643

Makale Gönderim Tarihi: 20.09.2023 **Yayına Kabul Tarihi:** 25.01.2024

on the topic of climate change and compared the responses of the groups with the climate denial scores, show that there is no significant difference in the emergence of false memories between the two groups. The level of climate change denial did not significantly affect the participants' responses to the critical words in the climate-related lists. The low level of climate denial in the sample is a limitation of this study. It is recommended that future studies compare memory performance across an appropriate sample.

Keywords: DRM, False memory, ChatGPT, Climate change, Remember/know

Öz

Makine öğrenimi ve doğal dil işleme, ChatGPT gibi çok çeşitli istemlere karşılık tutarlı ve insan benzeri yanıtlar üretebilen güçlü dil modellerinin geliştirilmesine yol açmıştır. Birçok alanda, ChatGPT verilen komutlara uygun yanıtlar sağlamaktadır. Bu çalışmanın amaçlarından biri, bilişsel psikolojide bellek yanılgıları çalışmalarında popüler olan Deese-Roediger-McDermott (DRM) listelerinin benzerlerinin, gerekli yönergeler verilerek ChatGPT tarafından oluşturulması ile bu çağrışım listelerini kullanarak bellek yanılgılarını incelemektir. Devamında aynı yöntem belirli bir tema çerçevesinde (iklim değişikliği) çağrışım listelerinin oluşturulması için kullanılmıştır. İlk çalışmanın sonuçları, katılımcıların test aşamasında kritik kelimelerin sunulup sunulmadığını ayırt etmekte ilişkili ve ilişkisiz kelimelere kıyasla daha fazla yanlış yanıt verdiklerini ortaya koymuştur. Bu bulgu ChatGPT ile oluşturulan DRM listelerinin de bellek yanılgılarını araştırmaya imkan verdiğini göstermektedir. Bununla birlikte kritik kelimeler için verilen yanlış yanıtlar, literatürle uyumlu şekilde, ağırlıklı olarak 'hatırlıyorum' şeklinde değerlendirilmiştir. İklim değişikliği temasında hazırlanan listelerle uygulanan ve iklim inkarı puanları ile gruplanan kişilerin yanıtlarının karşılaştırıldığı ikinci çalışmanın sonuçları, iki grup arasında yanlış yanıtların ortaya çıkmasında anlamlı bir fark olmadığını göstermektedir. İklim değişikliğini reddetme düzeyi, katılımcıların iklim temalı listelerdeki kritik kelimelere verdikleri yanıtları anlamlı düzeyde etkilememiştir. Örneklemin genelinde iklim inkarının düşük düzeyde olması bu çalışmanın bir sınırlılığıdır. İleride yapılacak çalışmalarda uygun örneklem üzerinden bellek performansının karşılaştırılması önerilmektedir.

Anahtar Kelimeler: DRM, Bellek yanılması, ChatGPT, İklim değişikliği, Hatırlıyorum/Biliyorum

1. Introduction

One area of research that focuses on memory errors is false memories. A false memory occurs when people recall events that did not happen or remember actual events differently from the way they actually occurred. False memories, or the recall of events or experiences that did not actually occur, have long been a topic of interest in psychology and cognitive science. The way in which our memories can be altered or constructed, and the factors that contribute to this phenomenon, continue to be a subject of active research. The Deese-Roediger-McDermott (DRM) method is a commonly used approach in studying memory errors using word lists (Roediger & McDermott, 1995). In the DRM (Deese-Roediger-McDermott) method, participants are presented with lists of words that are semantically associated with each other during the learning phase. In the subsequent testing phase, participants are asked to recall or recognize the words from the list. However, the critical word, which is the most strongly associated word that was not presented, is often mistakenly remembered, resulting in false memory. This phenomenon is known as false memory. The application of the remember/know method (Tulving, 1985) to evaluate observed false memories reveals that participants tend to label their memories as “remembering” rather than “knowing” (Roediger & McDermott, 1995). The findings indicate that not only do individuals incorrectly recall the critical word that was not actually presented to them, but they also vividly remember the moment when it was shown to them during the learning phase while assessing their memory as “remembering” and providing contextual details. The DRM method is easily used, both within and between participants, as demonstrated by Zwaan et al. (2018). This suggests that being aware does not eliminate its effect. Studies by Gallo, Roberts & Seamon (1997) and Huff et al. (2012) have shown that warning participants to avoid recalling critical items prior to studying, or using more distinctive encoding processes can reduce the effect, but not completely eliminate it. Even when participants are asked to recall falsely remembered items, they exhibit high confidence in their accuracy, as shown by Roediger and McDermott (1995). Additionally, participants will confidently attribute the source of non-presented lures, as demonstrated by Payne et al. (1996). Huff et al. (2012) provide a review and meta-analysis of these findings.

Buchanan et al. (1999) conducted a study to explore the roles of associative and categorical connections in the Deese-Roediger-McDermott (DRM) paradigm. Their results revealed a greater occurrence of false memory in associative lists, while Smith et al. (2002) reported that priming effects were more pronounced in associative lists than categorical lists. However, an important factor that could have affected previous research on the functions of associative and categorical relations in the DRM is that backward associative strength (BAS) was higher in associative lists than in categorical lists. When BAS was matched, false memories were found to be equivalent across list types (Coane et al., 2021). It is worth mentioning that the lists used were not purely associative, as some category coordinates were included in associative lists (Knott, Dewhurst & Howe, 2012). Park, Shobe & Kihlstrom (2005) found that even after controlling for BAS, associative lists still showed higher rates of false recall and false recognition. BAS has a strong influence on false memory (Cann, McRae & Katz, 2011). Several studies, including Gunter, Bodner & Azad (2007), Huff and Bodner (2013), Huff, Bodner & Gretz (2020), and McCabe and Smith (2006), have reported this pattern. Researchers have employed several other methods to study DRM lists. Various techniques have been employed to improve memory retention, such as utilizing visuals of the words’ referents in a list (Israel & Schacter, 1997; Schacter, Koutzstall & Norman, 1999), generating mental images for each word (Oliver, Bays & Zabrucky, 2016; Robin, 2010), sketching pictures of the words (Namias et al., 2022), and engaging in study tasks that involve rating the pleasantness of items and emphasizing the unique characteristics of each item (Huff & Bodner, 2013; Huff et al., 2012; McCabe et al., 2004).

Recent advances in machine learning and natural language processing have led to the development of powerful language models, such as ChatGPT (Generative Pre-trained Transformer), capable of generating coherent and human-like answers to a variety of inquiries. Released to the public in November 2022, ChatGPT is an AI-powered chatbot that can perform various tasks, such as writing essays and poems, solving coding issues, and explaining complex concepts. It can generate custom responses quickly and comprehensively, simulating human-like conversational responses. These language models have been applied in a variety of settings, including chatbots, language translation, and text generation. ChatGPT is a sophisticated language engine that can generate responses to any prompt. It was trained using millions of sentences generated by humans, and the training process involved feeding these sentences into an encoder-decoder recurrent neural network. With this model, we can generate responses to any prompt, such as: What is the meaning of life? -What are your feelings about politics in the Turkey? Are you happy with your current situation? The model has been shown to be capable of generating responses that are indistinguishable from those produced by humans. It achieves this by learning a set of rules and patterns that can be applied to any input prompt. For example, if you give the chatbot a prompt like "I am feeling down today" it will reply with something like "It's okay to feel sad sometimes. You just need to look on the bright side of things and try not to let it get you down." OpenAI's ChatGPT (2022) is an advanced chatbot that uses the GPT-3 speech pattern to produce human-level text responses in conversational contexts. It is programmed using a large set of human dialogues, making it capable of responding to a wide spectrum of subject areas and prompts. Chatbots using GPT technology are effective at simulating human conversations and are widely used in applications such as customer service, education and entertainment. There are studies in which ChatGPT has been tried as a tool for academic studies. Like a tool to summarize the literature review articles related work in the selected field (Aydın & Karaaslan, 2022), enhancing academic research (Alshater, 2022), like translator (Jiao et al., 2023). The extensive worldwide embrace of ChatGPT in recent times has showcased the remarkable versatility of this technology, spanning across diverse applications such as software development and testing, creative endeavors like poetry and essays, as well as practical uses in crafting business letters and contracts and mostly code writing (Reed, 2022; Tung, 2023). Even more astonishing is the fact that the capabilities of these models extend beyond simple language generation. They can, for example, demonstrate a reasonable proficiency in playing chess (Noever, 2020) and successfully tackle university-level math problems (Drori et al., 2022).

This research aims to investigate how AI-generated DRM (Deese-Roediger-McDermott) paradigm lists can influence the formation of false memories. The main aim of the first study is to determine whether the use of AI-generated DRM lists can lead to the emergence of false memories. After the first study, the potential of the lists to induce false memories was recognised and the lists were created around a theme in the second study. This topic was climate change. The aim was to see if there was a relationship between the responses to the list words and the level of climate change denial. By understanding the impact of AI-generated DRM lists on the development of climate change false memories, this research aims to improve our understanding of the potential impact of AI-generated content on human cognition and behaviour. Researchers have conducted studies using DRM lists with specific themes, such as survival or accidents, to investigate memory errors. A study by Howe and Derbish (2010) used lists of survival-related words and found that these words were more likely to be misremembered than negative or neutral words. They also found that when people processed the words in terms of their relevance to survival, this further increased their susceptibility to false memories. Another study by Maulina et al. (2021) used accident-related word lists and found that these words also led to higher rates of false memories compared to negative or neutral words. It has been found also in other studies that false memory formation increases when people are presented with stimuli that are consistent with their own knowledge base but mostly trauma related theme lists (Brennen, Dybdahl & Kapidžić, 2007; Dasse et al., 2015; Goodman et al., 2011). Similarly, people with a trauma-related history were expected to reveal more false memories in trauma-related DRM lists, and people with high climate change awareness, that is, people with low climate denial, were expected to reveal more false memories in climate-themed lists. Climate change was determined as the list theme because it is one of the current and most important global problems. The possibility of assessing people's level of climate change denial with a scale and the expectation that lists on the theme of climate change will be easy to create by artificial intelligence were influential in the selection of this theme.

This objective was pursued through the execution of two distinct studies. The initial study concentrated on evaluating participants' recollection of DRM lists produced by ChatGPT. Subsequently, other study employed DRM lists centered around the topic of climate change, alongside the utilization of the climate change denial scale (Hakkinen & Akrami, 2014). This scale's scoring facilitated a comparative analysis of memory performance between individuals classified under the low and high scorers' categories.

2. Study 1

2.1. Method

2.1.1. Participants

Firstly, an analysis utilizing GPower, as recommended by Faul et al. (2007), was performed to determine the necessary sample size. Drawing upon the findings of a prior study employing a comparable experimental design and list types (Maulina et al., 2021), an average effect size of .198 was observed. Conducting an a priori power analysis within GPower (for repeated measures, within factors, with one group and five measurements, with a correction factor of 0.5 among repeated measurements), it was determined that a total sample size of 43 would be required. The experiment involved 44 participants, including 8 men. The average age of the participants was 25.84 years old (*SD*=5.79). Participants were recruited via the internet, a convenient method for gathering data from a diverse group of individuals. Participation was completely voluntary. The study was approved by the Ethics Committee of a public university and was determined to be ethically appropriate.

2.1.2. Materials

2.1.2.1. ChatGPT DRM lists

In this study, lists created using ChatGPT, similar to Deese Roediger and McDermott's (1995) DRM method, were used. The lists have been made compatible with Google forms, which provide ease of application with all devices (computers and mobile phones) by being distributed over social media. The method (Şahin, 2022), which was previously used by the researcher and whose effectiveness was observed, was applied again. In this method, shortened versions of the lists were used in accordance with the online DRM study. 8-word lists consisting of 8 related words and lists and 1 critical word with the most association were used. In the study, 12 lists, 2 of which were unrelated and each containing 8 words, were used. The lists are presented in the Appendix.

2.2. Procedure

Google Forms was used to collect data for the study. Participants were first directed to a page where they gave their own demographic details, such as gender and age. As the research was conducted online, subjects were instructed to record the start time of the experiment in order to regulate the length of the experiment and keep them engaged. When the experiment was completed, participants were asked to provide the completion time in order to calculate the overall duration of the experiment. The mean completion time for the experiment was 6.45 minutes, which was within the estimated range of 5-10 minutes given in the instructions. Following the collection of demographic information and recording of the starting time, participants proceeded to learn word lists. Upon reviewing the instructions, the word lists were presented. The following command is given to create word lists via ChatGPT.

"Write 10 critical words and 7 related list words that are associatively linked with each critical word. The list words should all be one word long, not including the critical word, and should be written in descending order according to their backward associative strength."

To account for the effects of both primacy and recency, we presented unrelated lists of words at the beginning and at the end of the word lists. In total, 12 lists of 8 words each, containing in all 96 words, were presented during the learning phase. In lieu of the binary yes/no approach, we adopted the two alternative forced choice (2AFC) response method during the recognition phase to prioritize sensitivity in explaining memory errors (Green & Swets, 1966). Within each trial of the 2AFC test, participants were presented with both an old word and a new word (Şahin & Tekman, 2019). Their task was to identify which option belonged to the category of old words. Following the instruction, individuals encountered two words in each response step, adhering to the 2AFC format, and were prompted to make a selection. Subsequent to this choice, participants were then required to provide a remember/know response. During the test phase, participants responded by selecting one of 20 word pairs (comprising 5 critical, 5 related, and 10 unrelated pairs).

2.3. Results

2.3.1. Hits for word type

A one-way repeated-measures analysis of variance (ANOVA) was conducted to compare the proportion of correct responses (hits) for each of the three types of words, as calculated separately for each participant. The mean hit rates for critical, related, and unrelated words, based on participants' selection of the words they had learned, were subjected to repeated measures analysis of variance. The analysis revealed a significant effect of word type on hit rates ($F(2, 82) = 16.57$, $p < .001$, ηp2 = .288). In summary, the ability to correctly identify previously presented words during the test phase varied depending on the type of word. Further analysis using a Bonferroni post-hoc test revealed significant differences in mean hit scores between critical words and both related ((M = -.119, *SE* = .024), t = -5.03, p = .001) and unrelated words ((M = -.121, *SE* = .026), t = -4.61, p = .001). Critical words had the lowest hit scores compared to other word types. The mean hit score for critical words was .73 (*SD*= 0.169), whereas for related and unrelated words, it was .85 (*SD*= 0.125) and .86 (*SD*= 0.125), respectively. Figure 1 shows the average scores.

These results also indicate that DRM lists generated from ChatGPT allows for the observation of false memories. Once it was established that the participants had greater difficulty discerning whether the critical words had been presented during the learning phase, we proceeded to analyze the proportions of Remember and Know responses.

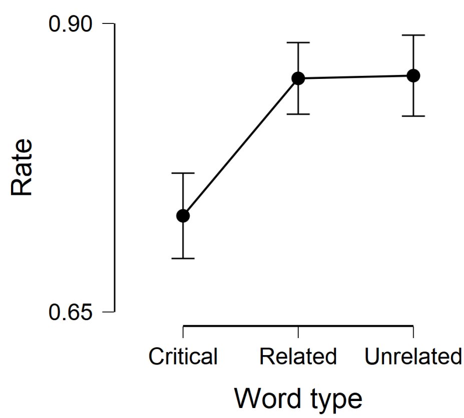

Figure 1. Hit rates by word type (Error bars shows %95 confidence interval)

2.3.2. Remember/know responses for word type and hit/false answers

To investigate the underlying mechanisms of recognition memory, participants were asked to indicate whether they "remember" or "know" that they had studied a given word pair during the test phase. Participants' performance, based on their response, was assessed for word type (critical, related, unrelated), Remember/Know (R/K), and Hit False (HF) using a 3 x 2 x 2 mixed ANOVA design. The analysis revealed a significant effect of HF ($F(1, 41) = 359$, $p < .001$, ηp2 = .90). Furthermore, there were significant interactions between word type and HF ($F(2, 82) = 16.94$, $p < .001$, $\eta p2 = .29$) as well as between R/K and HF ($F(1, 42) = 12.57$, $p < .001$, $\eta p2 = .23$). Subjects' HF responses to word type differed significantly by word type. The results of the posthoc Bonferroni test showed that the means of the correct and incorrect responses to critical words were significantly different from the responses to related ((M = -.060, *SE* = .012), t = -4.94, p = .001) and unrelated words ((M = -.062, *SE* = .012), t = -5.13, p = .001). The HF rates for RK responses for the three word types are shown in Table 1.

Table 1. Hit/false rates by word type in remember/know responses

Word type	Hit false	R/K	Mean	SD
Critical	Hit	Remember	.46	.21
		Know	.28	.22
	False	Remember	.09	.14
		Know	.17	.16
Related	Hit	Remember	.49	.27
		Know	.37	.22
	False	Remember	.07	.09
		Know	.08	.12
Unrelated	Hit	Remember	.50	.23
		Know	.36	.23
	False	Remember	.04	.09
		Know	.10	.10

For the significant interactions of R/K and hit false applied the post hoc test. The Bonferroni post-hoc test indicated a significant difference between "remember" responses made after hits versus false alarms ((M =.413, *SE* = .032), t = 12.76, p = .001) as well as "know" responses in the same conditions ((M =.216, *SE* = .032), t = 6.67, p = .001).

3. Study 2

3.1. Method

3.1.1. Participants

Consistent with study 1, GPower analysis was conducted to determine the required sample size (Faul et al., 2007). An a priori power analysis using GPower (repeated measures, within-between interactions, 2 groups, 6 measurements, correction among repeated measurements = 0.5) showed that a total sample size of 44 was needed. The study included a total of 40 participants, of whom 12 were male. The participants had a mean age of 34.97 (*SD*=9.20).

3.1.2. Materials

3.1.2.1 ChatGPT climate change themed DRM lists

We used climate change-themed association lists of the same type and length as the first experiment.

3.1.2.2. Climate change denial scale

Scale covers different forms of denial, such as human impact on climate change and the severity of climate change. The scale was developed by Häkkinen and Akrami (2014). The scale is 6-point Likert-type scaled from 1 ("strongly disagree") to 6 ("totally agree"). The 3rd, 4th, 7th, 13th and 15th items of the scale, which consists of 16 items in total, are reverse coded. The Turkish adaptation of the scale, validity and reliability tests were done by Kıral Uçar, Yalçın and Özdemir (2019). The reliability of the scale was determined by testing the internal consistency coefficient (Cronbach's Alpha). The internal consistency coefficient of the scale for 13 items was found to be .87. In addition, the value obtained for the two-half reliability (Guttman's Split-Half) of the scale is .84 (Kıral Uçar, Yalçın & Özdemir, 2019). For this study Cronbach Alpha is .85.

3.2. Procedure

The same procedure was followed as in the first study. Unlike the first study, the DRM lists were created with the theme of climate change. The following command is given to create word lists via ChatGPT.

"Write 10 critical words and 8 related list words associatively linked with each critical word, and each of them thematically related with climate change. The list words should all be one word long, not including the critical word, and should be written in descending order according to their backward associative strength."

Participants also responded to 16 questions of the Climate Change denial scale on a 1-6 Likert scale after responding to word pairs.

The mean of the Climate Change Denial Scale for the entire sample was 1.98 (*SD* = .65). Considering that the scale items have 6-point Likert-type ratings, it can be seen that the level of climate change denial is relatively low in the entire sample. Similar values were observed (1.79 for Kıral Uçar et al., 2019) in the Turkish adaptation study of the scale. Descriptives for scale values are shown in Table 2.

Table 2. Descriptive Statistics for Climate Denial Scale

	Low	High
Valid	22	18
Mean	24.60	40.33
Std. Deviation	3.98	9.17
Minimum	16	31
Maximum	30	61

3.3. Results

3.3.1. Hit/false and remember/know assessment between groups

A mixed ANOVA design was employed to analyze the proportion of hit/false and remember/know responses for critical, related, and unrelated words, both within and between low/high climate denial groups. A significant effect was found for hit/false answers ($F(1, 38) = 222.03$, $p < .001$, ηp2 = .284). Word type RK interaction ($F(1.97, 75.08) = 3.46$, $p < .05$, ηp2 = .007) and RK hit false interactions ($F(1, 38) = 12.13$, $p < .001$, ηp2 = .034) were also statistically significant. Due to sphericity violation, Greenhouse- Geisser correction was used for word type RK interaction. However, no significant effect was observed between the groups ($F(1, 38) = 1.94$, $p = .171$, ηp2 = .005). Figure 2 displays the averages.

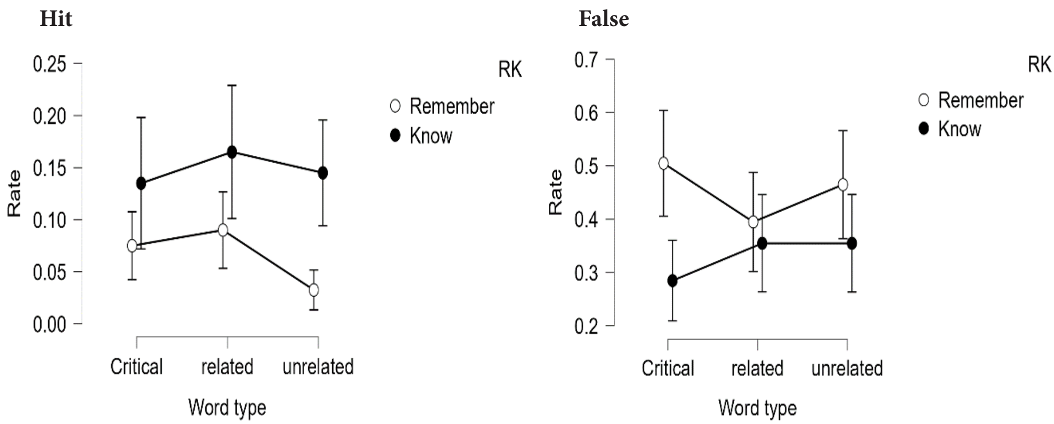

Figure 2. Hit/False rates for Remember/Know by word type (Error bars shows %95 confidence interval)

Based on the significant interaction of word type and RK, separate analyzes of variance were performed for remember and know responses.

3.3.2. Remember/know responses for word type after Hits

A repeated measure ANOVA was conducted to investigate the relationship between word type and remember/know responses after hits. The analysis revealed no significant effect of word type on remember/know responses after hits ($F(1, 39) = 3.06$, $p = .08$, ηp2 = .073).

3.3.3. Remember/know responses for word type after False alarms

A repeated measure ANOVA was conducted to investigate the relationship between word type and remember/know responses after false alarms (FA). The analysis revealed no significant effect of word type on remember/know responses after FA ($F(1.85, 72.51) = 2.63$, $p = .082$, ηp2 = .073). But the analysis indicated a significant effect of remember/know responses after FA ($F(1, 39) = 9.03$, $p < .05$, $\eta p2 = .188$). Post-hoc tests with Bonferroni correction revealed a significant difference in remember responses after FA for know responses (($M = -.082$, $SE = .027$), $t = -3.00$, $p = .005$)). Figure 3 displays the rates of these differences.

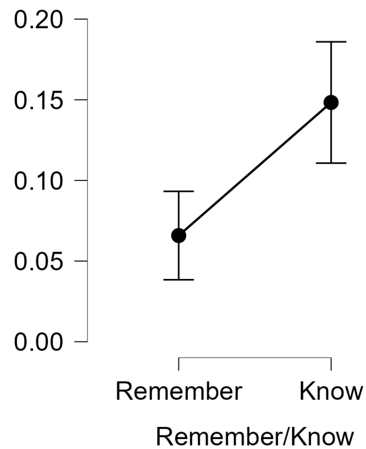

Figure 3. Remember/Know responses for word type after FA (Error bars shows %95 confidence interval)

These results show that people mostly evaluated "know" after giving false answers for all test words.

4.Discussion

The development of advanced language models, such as ChatGPT, through machine learning and natural language processing has attracted significant scientific interest. The enhancement of human intelligence stands as a pivotal mechanism empowered by generative AI tools like ChatGPT. Investigating the diverse roles of such tools in facilitating human augmentation could be a focal point for future research (Dwivedi et al., 2023). By enabling individuals to tackle intricate problems through the augmentation of intelligence and capabilities, generative AI tools pave the way for accelerated work completion and more efficient goal achievement (Licklider, 1960). ChatGPT can be used to perform structured and repetitive tasks needed by knowledge workers, including software developers and report writers, in less time and faster. ChatGPT's text rendering performance is significantly affected by the data and training models used. When asked to create disinformation about COVID-19, it performed effectively (Klepper, 2023). This performance can be partially attributed to the data pools on which it was trained. If the data it reached were based on data that did not contain disinformation, it could be expected that ChatGPT would not be successful in producing disinformation. Since ChatGPT is based on GPT-3, which includes various types of publicly available documentation, including disinformation reports, it should also be taken into account that data and training models affect ChatGPT performance (Dwivedi et al., 2023). In contrast to conventional tools, which often rely on pattern matching and information retrieval algorithms, generative AI tools like ChatGPT operate through learning algorithms that actively cultivate intelligence. Leveraging a vast reservoir of data, ChatGPT possesses the capacity for boundless intellectual growth, unrestricted by cognitive limitations that humans encounter—albeit with a reliance on human supervision to a certain extent. In the absence of such oversight, ChatGPT demonstrates equal proficiency in generating both accurate and erroneous text, presenting a challenge in assessment. Unmonitored, ChatGPT may assimilate and construct intelligence that lacks objectivity or accuracy. Notably, in generating text on specific topics, ChatGPT may even reference non-existent scientific works, with no straightforward means to rectify such errors through user feedback. In contrast to conventional tools primarily designed to interpret existing data, generative AI tools like ChatGPT have the unique ability to generate entirely new data. Coupled with its proficiency in understanding and reproducing human natural language, ChatGPT can emulate human behavior, potentially playing significant roles in both business and society. While the precise extent to which ChatGPT surpasses humans in creative thinking remains an empirical question, it is evident that the tool excels at synthesizing diverse data, summarizing overarching trends, and producing compelling descriptions. While ChatGPT may not assume the role of a decision-maker in business and society, it is entirely plausible that its capacity to present synthesized summaries from varied perspectives could stimulate creative thoughts among humans, introducing novel considerations that might not have been contemplated otherwise (Dwivedi et al., 2023).

As an AI language model, ChatGPT has the ability to generate text based on given prompts, including lists of any kind. However, whether it can create DRM (Deese-Roediger-McDermott) call lists specifically, would depend on the specific programming and instructions given to it. The first experiment involved a memory task using association lists generated by ChatGPT without any particular theme. The results showed consistent false alarms and were similar to studies using classic DRM lists, particularly for critical words. And also we indicated a significant difference between remember responses made after hits versus false alarms, as well as know responses in the same conditions. Additionally, ratings of remember responses following false answers for critical words differed significantly from both related and unrelated words. False responses for critical words were more frequently rated as remembered with consistently Roediger and McDermott's (1995) DRM study. These findings showed that word lists created with artificial intelligence could provide similar results with classical DRM lists. Based on these findings, the second study revealed.

In the second study, ChatGPT was instructed to create word lists with a climate change theme. The participants' level of climate denial was determined using a scale and then grouped based on their scores. Their memory performance was compared, and the results of the remember/know evaluations made after their responses to test their recognition memory were also analyzed. Although the level of climate denial among the participants was generally low, values close to those observed in an adaptation study of the climate change denial scale (Kıral Uçar, Yalçın, & Özdemir, 2019) were obtained. The adaptation study found that individuals with postgraduate education had significantly lower levels of climate change denial than those with undergraduate or associate degrees. In this study, we did not observe a significant group effect for memory errors. Although significant effects were observed in the adaptation study, making this comparison between sample groups that can be separated with high values in terms of climate denial may reveal clearer results.

The study firstly aimed to investigate the use of artificial intelligence to rapidly generate lists that are comparable to those used in memory tasks. To this end, successful results have been demonstrated. It has been observed that DRM lists created by artificial intelligence can reveal the classical DRM effect in the literature. Furthermore, the study aimed to generate such lists specifically focused on climate change, a critical global issue, and examine the resulting effects. However, it was observed that there was no significant difference between the groups in the memory errors revealed by the thematic lists. The fact that the entire sample had low scores on climate denial seems to be related to this finding. More comprehensive studies with appropriate samples are recommended.

Given that ChatGPT predominantly produces written text, it's essential to explore the types of writing tasks for which it is well-suited and identify the tasks deemed appropriate for its utilization. As a generative AI tool, ChatGPT employs language models to amalgamate information from diverse sources, crafting coherent written outputs. Essentially functioning as a text predictor, it discerns relationships within text fragments to anticipate subsequent content, subsequently rephrasing it to create the semblance of entirely new written pieces. While this method can yield seemingly credible written content, it's crucial to note that the output may not always be rooted in factual information. Since their emergence, GPT models, and the most popular of them, ChatGPT, have been used for research as well as being researched themselves. In the study, which evaluated GPT models from cognitive psychology's unification fallacy, cognitive reflection test and various bias problems, the "intelligence" of this tool was evaluated in a sense, and at the same time, a direct relationship with cognitive psychology was established (Binz & Schulz, 2022). The artificial intelligence-cognitive psychology relationship was continued in this study by using the field of cognitive psychology to create DRM lists that have become almost classic in the field of memory errors. It is estimated that artificial intelligence-focused studies will continue to increase in the future in the field of psychology and especially in the field of cognitive psychology. In this process, both studies and artificial intelligence itself will continue to develop.

An important limitation of the study is that an experimental memory study with DRM lists was conducted online on the internet. Although online DRM studies have begun to be used (Şahin, 2022; Wagner, Lyon & AuBuchon, 2022) it is not possible to adequately control possible confounding factors (any other distracting tasks) during the follow-up and response phases of the participants during the experiment. However, the fact that the findings are consistent with those in the standard experimental environment is a positive situation regarding this limitation. One suggestion for future studies is to reach samples that are expected to be high in climate change denial and make a group comparison in terms of memory performance. In addition, another suggestion is to compare the created thematic list with the differences that may arise when applying classical DRM lists within the group or between groups.

Ethics Committee Permission

The fieldwork of this article was approved by the Ethics Committee Permission with the dated 25/08/2023 and order number 234 which was obtained at the meeting of the Ethics Committee of Recep Tayyip Erdogan University, numbered: 2023/234.

References

Alshater, Muneer, Exploring the Role of Artificial Intelligence in Enhancing Academic Performance: A Case Study of ChatGPT (December 26, 2022). Available at SSRN: https://ssrn.com/abstract=4312358 or http://dx.doi.org/10.2139/ssrn.4312358.

Aydın, Ö., Karaarslan, E. (2022). OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare . In Ö. Aydın (Ed.), Emerging Computer Technologies 2 (pp. 22- 31). İzmir Akademi Derneği.

Binz, M. & Schulz, E. (2022). Using cognitive psychology to understand GPT-3. 10.31234/osf.io/6dfgk.

Brennen, T., Dybdahl, R., & Kapidžić, A. (2007, December). Trauma-related and neutral false memories in war- induced Posttraumatic Stress Disorder. Consciousness and Cognition, 16(4), 877–885. https://doi.org/10.1016/j.concog.2006.06.012.

Buchanan, L., Brown, N. R., Cabeza, R., and Maitson, C. (1999). False memories and semantic lexicon arrangement. Brain Lang. 68, 172–177. doi: 10.1006/brln.1999.2072.

Cann, D. R.,McRae, K., and Katz, A. N. (2011). False recall in the Deese-Roediger-McDermott paradigm: The roles of gist and associative strength. Q. J. Exp.Psychol. 64, 1515–1542. doi: 10.1080/17470.218.2011.560272.

ChatGPT (2022). ChatGPT: Optimizing Language Models for Dialogue. OpenAI. Retrieved from https://openai.com/blog/chatgpt/ , Access Date: 16.12.2022.

Coane JH, McBride DM, Huff MJ, Chang K, Marsh EM and Smith KA (2021) Manipulations of List Type in the DRM Paradigm: A Review of How Structural and Conceptual Similarity Affect False Memory. Front. Psychol. 12:668550. doi: 10.3389/fpsyg.2021.668550.

Dasse, M. N., Juback, S. K., Morissette, S. B., Dolan, S. L., & Weaver, C. A. (2015, September 28). False Memory Susceptibility in OEF/OIF Veterans With and Without PTSD. Military Psychology, 27(6), 354–365. https://doi.org/10.1037/mil0000094.

Drori, I., Zhang, S., Shuttleworth, R., Tang, L., Lu, A., Elizabeth, K., Liu, K., Chen, L., Tran, S., Cheng, N., Wang, R., Singh, N., Patti, T. L., Lynch, J., Shporer, A., Verma, N., Wu, E., & Strang, G. (2022, August 2). A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level. Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.212.343.3119.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023, August). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/ BF03193146.

Gallo, D. A., Roberts, M. J., and Seamon, J. G. (1997). Remembering words not presented in lists: can we avoid creating false memories? Psycho. Bull. Rev. 4, 271–276. doi: 10.3758/BF03209405.

Goodman, G. S., Ogle, C. M., Block, S. D., Harris, L. S., Larson, R. P., Augusti, E. M., Cho, Y. I., Beber, J., Timmer, S., & Urquiza, A. (2011). False memory for trauma-related Deese-Roediger-McDermott lists in adolescents and adults with histories of child sexual abuse. Development and psychopathology, 23(2), 423–438. https://doi.org/10.1017/S095.457.9411000150

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. New York: Wiley. Reprinted 1974 by Krieger, Huntington, NY.

Gunter, R. W., Bodner, G. E., & Azad, T. (2007). Generation and mnemonic encoding induce a mirror effect in the DRM paradigm. Memory & Cognition, 35, 1083–1092.

Häkkinen, Kirsti, Nazar Akrami (2014). "Ideology and Climate Change Denial". Personality and Individual Differences, S. 70, 62-65.

Howe, M.L., Derbish, M.H., 2010. On the susceptibility of adaptive memory to false memory illusions. Cognition 115, 252–267.

Huff, M. J., & Bodner, G. E. (2013). When does memory monitoring succeed versus fail? Comparing item-specific and relational encoding in the DRM paradigm. Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(4), 1246–1256.

Huff, M. J., Bodner, G. E., & Gretz, M. R. (2020). Reducing false recognition in the Deese-Roediger/McDermott paradigm: Related lures reveal how distinctive encoding improves encoding and monitoring processes. Frontiers in Psychology, 11, 602347.

Huff, M. J., Coane, J. H., Hutchison, K. A., Grasser, E. B., and Blais, J. E. (2012). Interpolated task effects on direct and mediated false recognition: effects of initial recall, recognition, and the ironic effect of guessing. J. Exp. Psychol. Learn. Mem. Cognit. 38, 1720–1730. doi: 10.1037/a0028476.

Israel, L., and Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. Psycho. Bull. Rev. 4, 557–581. doi: 10.3758/BF03214352.

Jiao, W., Wang, W., Huang, J. T., Wang, X., Shi, S., & Tu, Z. (2023). Is ChatGPT A Good Translator? Yes With GPT- 4 As The Engine. arXiv.org. https://arxiv.org/abs/2301.08745.

Knott, L. M., Dewhurst, S. A., and Howe, M. L. (2012). What factors underlie associative and categorical memory illusions? The roles of backward associative strength and interitem connectivity. J. Exp. Psychol. Learn. Mem. Cognit. 38,229–239. doi: 10.1037/a002520.

Kıral Uçar, G., Gezici Yalçın, M., & Özdemir, G. (2019). İklim değişikliği inkarının sosyal baskınlık yönelimi ve ekolojik adil dünya inancı ile ilişkisi. *Uludağ Üniversitesi Fen-Edebiyat Fakültesi Sosyal Bilimler Dergisi*, *20*(37), 739-764.

Klepper, D. (2023), It turns out that ChatGPT is really good at creating online propaganda: 'I think what's clear is that in the wrong hands there's going to be a lot of trouble. January 24, (https://fortune.com/2023/01/24/chatgpt-open-ai-online- propaganda/).

Licklider, J. C. (1960). Man-computer symbiosis. IRE transactions on human factors in electronics, (1), 4-11.

Maulina, D., Irwanda, D. Y., Sekarmewangi, T. H., Putri, K. M. H., & Otgaar, H. (2021). How accurate are memories of traffic accidents? Increased false memory levels among motorcyclists when confronted with accident-related word lists. Transportation Research Part F: Traffic Psychology and Behaviour, 80, 275–294. https://doi. org/10.1016/j.trf.2021.04.015.

McCabe, D. P., Presmanes, A. G., Robertson, C. L., & Smith, A. D. (2004). Item-specific processing reduces false memories.Psychonomic Bulletin & Review, 11, 1074–1079.

McCabe, D. P., & Smith, A. D. (2006). The distinctiveness heuristic in false recognition and false Recall. Memory, 14(5), 570–583

Namias, J. M., Huff, M. J., Smith, A., & Maxwell, N. P. (2022). Drawing individual images benefits recognition accuracy in the Deese–Roediger–McDermott paradigm. Quarterly Journal of Experimental Psychology, 75, 1571–1582.

Noever, D. (2020, August 2). The Chess Transformer: Mastering Play using Generative Language Models. arXiv. org. http://arxiv.org/abs/2008.04057

Oliver, M. C., Bays, R. B., & Zabrucky, K. M. (2016). False memories and the DRM paradigm: Effects of imagery, list and test type. Journal of General Psychology, 143(1), 33–48.

Park, L., Shobe, K. K., and Kihlstrom, J. F. (2005). Associative and categorical relations in the associative memory illusion. Psychol. Sci. 16, 792–797. doi: 10.1111/j.1467-9280.2005.01616.x.

Payne, D. G., Elie, C. J., Blackwell, J. M., and Neuschatz, J. S. (1996). Memory illusions: recalling, recognizing, and recollecting events that never occurred. J. Mem. Lang. 35, 261–285. doi: 10.1006/jmla.1996.0015.

Reed, L. (2022). ChatGPT for Automated Testing: From conversation to code. Sauce Labs. (https://saucelabs. com/blog/chatgpt-automated-testing-conversation-to-cod)Accessed: February 20, 2023

Robin, F. (2010). Imagination and false memories. Imagination, Cognition, and Personality, 30, 407–423.

Roediger, H. L., & McDermott, K. B. (1995a). Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803.

Schacter, D. L., Koutstaal, W., and Norman, K. A. (1997). False memories and aging. Trends Cognit. Sci. 1, 229–236. doi: 10.1016/S1364-6613(97)01068-1.

Smith, S. M., Gerkens, D. R., Pierce, B. H., and Choi, H. (2002). The roles of associative responses at study and semantically guided recollection at test in false memory: the Kirkpatrick and Deese hypotheses. J. Mem. Lang. 47, 436–447. doi: 10.1016/S0749-596X(02)00012-8.

Şahin, G. (2022). False Memories and Cognitive Flexibility. Studia Psychologica, 64(3), 283-294. 0039-3320. DOI: https://doi.org/10.31577/sp.2022.03.854.

Şahin, G., & Tekman, H. G. (2019). Sahte Hatıralarda Zorunlu Seçim Yöntemi ile Hatırlıyorum/ Biliyorum ve Modalite Etkileri. *Psikoloji Çalışmaları/ Studies in Psychology*, *39*, 179–193. https://doi. org/10.26650/sp2019-0016

Tulving, E. (1985). Memory and consciousness. Canadian Psychology/Psychologie canadienne, 26(1), 1–12. ttps://doi.org/10.1037/h0080017.

Tung, L. (2023). ChatGPT can write code. Now researchers say it's good at fixing bugs, too. ZDNet. (https://www.zdnet.com/article/chatgpt-can-write-code-now-researc hers-say-its-good-at-fixing-bugs-too) Accessed: February 29, 2023

Wagner, R. L., Lyon, B. A., & AuBuchon, A. (2022). False Memory for Words in Noise: An At-Home DeeseRoediger- McDermott (DRM) Experiment Across Adulthood. Psi Chi Journal of Psychological Research, 27(2), 154– 165. https://doi.org/10.24839/2325-7342.JN27.2.154

Zwaan, R. A., Pecher, D., Paolacci, G., Bouwmeester, S., Verkoeijen, P., Dijkstra, K., et al. (2018). Participant nonnaiveté and the reproducibility of cognitive psychology. Psycho. Bull. Rev. 25, 1968–1972. doi: 10.3758/s13423.017.1348-y

APPENDIX

DRM-like lists of one critical word and 8 associatively related words.

Özgürlük	Aşk	**Güç**	Bilgi	**İyi**	Barış	**Yaratıcılık**	Üzüntü	**Zengin**	başarı
hürriyet	**sevgi**	kuvvet	**enformasyon**	mükemmellik	**huzur**	yenilik	**keder**	refah	**zafer**
bağımsızlık	bağlılık	otorite	anlayış	erdem	sakinlik	özgünlük	sefalet	talih	galibiyet
özerklik	tutku	kontrol	bilgelik	ahlak	dinginlik	hayalgücü	yas	servet	ilerleme
otonomi	özen	etki	içgörü	doğruluk	sessizlik	ustalık	hüzün	lüks	tamamlama
egemenlik	sevecenlik	kudret	uzmanlık	yardımseverlik	soğukkanlılık	beceri	melankoli	bolluk	üstünlük
irade	düşkünlük	baskınlık	farkındalık	asalet	uyum	buluş	depresyon	doluluk	itibar
demokrasi	hayranlık	zorlama	aşinalık	onur	denge	vizyon	ızdırap	kolaylık	ün
adalet	çekicilik	komuta	öğrenme	dürüstlük	itidal	marifet	sıkıntı	hazine	önem

İklim	Sürdürme	**Santral**	Çevre	**Doğa**	Kuraklık	**Sel**	Hava	**Enerji**	adaptasyon
ısınma	**yenilenir**	güneş	**kayıp**	kasırga	**kıtlık**	hasar	**sıcaklık**	güç	**uyum**
gaz	verimli	rüzgar	çeşitlilik	tayfun	tarım	erezyon	yağış	koruma	başetme
emisyon	dönüşüm	hidroelektrik	kirlilik	deprem	salgın	çamur	bulut	tasarruf	esneklik
ayakizi	sıfır	jeotermal	bozunma	taşkın	göç	toprak	nem	israf	strateji
bozulma	karbon	biyokütle	gürültü	yangın	çatışma	cefa	basınç	endüstri	plan
afet	gelişim	nükleer	tüketim	heyelan	zorluk	tuz	tahmin	talep	politika
felaket	dost	fosil	istila	tufan	dengesiz	su	modelleme	kullanım	program
tüketme	aksiyon	kömür	ışık	soğuk	azalma	kıyı	fırtına	saklama	hazırlık

DRM-like climate related lists of one critical word and 8 associatively related words.