Research Article

AI vs. Human in Screenwriting: Is AI the Future Screenwriter?

Kemal Çelik

İstanbul Medipol University, Department of Radio Television Cinema, İstanbul, Türkiye, kemal.celik@medipol.edu.tr

Abstract: This research paper investigates the challenge of discerning between content generated by humans and AI language models across various screenwriting elements. The study presents loglines, treatments, synopses, and dialogue examples to a group of 24 participants. These script elements are created by one human scriptwriter and an AI model. The participants are persons who have taken scenario courses, selected by purposive sampling method. The AI model used to generate the script content is Bing Chat, which is powered by ChatGPT 4, a conversational neural network known for its ability to produce natural and coherent text. Both AI-generated and human-created screenplay elements were provided to the participants, who were then asked questions related to them. The method of presenting both human and AI-generated screenplay options to each participant was chosen because it allows for a direct comparison and an accurate measure of the participants' ability to distinguish between the two. The structured interview type, which is also known as standard interview and researcher-managed interview, was used in this study. The average accuracy rate for all the script elements was 52.7%, which means that the participants had a moderate level of success in identifying the AI-generated and human-generated script elements. The study shows that artificial intelligence can produce script elements that are comparable to, and sometimes even preferred over, human-generated ones.

Keywords: Artificial Intelligence, Script, Scriptwriter, Future of Cinema, ChatGPT

1. Introduction

Artificial intelligence (AI) is one of the most current topics in various fields of study and practice. It is considered a turning point in human history for its achievements and potentials. As Mijwel (2015) summarized, AI is the technology that simulates human intelligence processes by computer systems, such as learning, reasoning, problem-solving, speech recognition, and planning. The history of AI dates back to ancient times, when myths and legends about artificial creatures and automata were prevalent. However, the modern AI emerged in the mid-20th century, with the development of digital computers and mathematical models of computation. Some of the pioneers and visionaries of AI include Alan Turing, John von Neumann, Claude Shannon, Warren McCulloch, Walter Pitts, and John McCarthy. AI has gone through several phases of development, such as symbolic AI, cybernetic AI, expert systems, neural networks, and machine learning. AI has also been applied to various fields and domains, such as game playing, natural language processing, computer vision, robotics, and healthcare. The current state of AI is characterized by rapid progress and innovation, driven by advances in hardware, software, data, and algorithms. AI also poses many challenges and opportunities for society, such as ethical, legal, social, and economic issues. AI is expected to continue to evolve and transform human life in the future.

Similarly, Artificial intelligence (AI) plays an increasingly important role in content creation today. AI can produce content in different forms, such as text, audio, image, and video, and contribute to the creative process of humans. However, there are also some problems and debates regarding the quality, consistency, and originality of the content generated by AI.

The influence of artificial intelligence has grown so significantly that screenwriters are now calling for limitations on its application. During the recent Writers Guild of America (WGA) strike, the union advocated for regulations governing the use of AI in generating source material and in writing or rewriting literary content. Their objective was to restrict the use of AI in film and television scripts. In

response, the Alliance of Motion Picture and Television Producers (AMPTP) proposed holding "annual meetings to discuss advancements in technology" (Shah, 2023). Although the WGA's contract does not prohibit the use of AI tools in the writing process, it establishes safeguards to ensure that this emerging technology remains under the control of human workers. These new guidelines are designed to prevent several scenarios that writers had previously feared (Anguiano & Beckett, 2023).

One of the most competent AI models that is at the center of these debates is ChatGPT. As Marr (2023) explained in detail, ChatGPT is an AI chatbot developed on top of OpenAI's large language models like GPT-4. OpenAI introduced an early demo of ChatGPT on November 30, 2022, and it gained widespread attention on social media as users shared its capabilities in various tasks, from travel planning to coding. Within five days, it attracted over one million users. ChatGPT has influenced various industries including customer service, education, content creation, businesses, healthcare, and entertainment. This model, which is the GPT-4 version of AI accessed through Bing chat, was used for AI screenwriter production for this study.

1.1. Research objectives

The primary objective of this paper is to test whether AI-generated and human-crafted screenplays can be distinguished by conducting standardized open-ended interviews with participants involving various screenwriting elements, including loglines, treatments, synopses, and dialogue examples. Other objective is to discuss the implications of the findings for screenwriters, educators, and the creative industry. The findings of this research can provide insights into the strengths and weaknesses of AI models in generating screenwriting content, as well as the challenges and opportunities for human screenwriters in adapting to the AI era. The findings can also inform educators on how to design effective curricula and pedagogies for teaching screenwriting skills in the context of AI. Thus, this question will answer whether artificial intelligence can generate scripts as appreciated or even more loved than those produced by human screenwriters.

1.2. Literature review

There is a vast literature on the creativity competition between artificial intelligence and humans. The Turing test can be cited as one of the earliest studies in this field. Turing proposes the idea of the learning machines that can simulate a child's mind, which is easier to program than an adult's mind. He says that a machine can learn from both physical and cognitive experiences, even if they seem random. (Turing, 1950, pp. 434-456). This debate is already the most prominent discussion in the field: Can machines act or think like humans?

So, the debate on whether artificial intelligence can think like humans is ongoing. McCarthy (2007, pp. 2-5) suggests that intelligence is not a simple concept that can be answered with a yes or no for machines. He says that intelligence involves different mechanisms that computers can do some of them, but not others.

Ray Kurzweil (2005, pp. 127-128) explores the concept of singularity, which is the point where a machine can match or exceed human intelligence. This discussion is addressed by a dissertation by Kurt on artistic creativity in artificial intelligence (Kurt, 2018). The topic of the dissertation is whether artificial intelligence can produce artistic works. This paper explores the intersection of AI, creativity, and artistry, challenging traditional definitions of art and offering a conceptual framework for understanding AI art as a legitimate and autonomous art genre. It highlights the distinctiveness of AI-generated art, characterized by its lack of emotional self-expression and the resulting aura of uniqueness.

One of the inspiring studies that brings together the screenwriter and artificial intelligence directly is a program described in the work titled "Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals" (Mirowski, Mathewson, Pittman, & Evans, 2023). This

work is about Dramatron, an interactive co-authorship script writing tool that leverages large language models such as Chinchilla. The authors conclude that Dramatron can be used as a co-creative writing tool that can help human writers generate scripts and screenplays from a single log line. They also highlight the benefits and limitations of using large language models for creative writing.

As seen in all these studies, the discussions focus on the creativity aspect of artificial intelligence. Screenwriting is a profession that requires creativity directly, so the question of how competent artificial intelligence can be in this regard is at the center of the debate. Moreover, the importance of the script in film production makes the issue even more significant.

As Li detailed (2022, pp. 1-2), script writing is a vital and complex part of film pre-production. Scripts need to be original and expressive to appeal to the audience. This requires a lot of creativity and skill from the creator. Artificial intelligence can help with script writing by accessing and analyzing online information, selecting stories and references, and avoiding duplication. AI can also write scripts faster than humans. Script-writing AI already exists, such as the software developed by Andy Herd in 2016 using Google's tensorflow, which generated a new episode of Friends. Another example is Benjamin, an AI program that was a finalist in the sci-fi London 48-hour challenge in 2016 with his script sun-sprint, and won a special mention in 2018 for his zone out.

Indeed, the central question that emerges from all these examples is whether artificial intelligence can replace human screenwriters. The article "Analysis on Whether Artificial Intelligence Can Replace Human Screenwriters" by Song (2022) discusses the same topic. This paper explores the impact of artificial intelligence (AI) on scriptwriting and literary creation. The paper concludes that AI and human creativity can coexist, with AI supporting and complementing the work of scriptwriters and other literary creators.

Although Song's claim that AI cannot replace human in screenwriting is somewhat theoretical, it raises an important debate on the copyright issues arising from AI in screenwriting. Another paper, authored by Kavitha L. and published in July 2023, delves into the intricate intersection of AI technologies, scriptwriting, content generation, and copyright issues. The study investigates the potential pitfalls of AI-generated content resembling or infringing upon copyrighted materials, emphasizing the importance of creators and organizations navigating these challenges while adhering to copyright laws.

Another paper (Wilson & Anderson, 2022) explores the confluence of artificial intelligence (AI) and creativity, focusing on the creation of a Gothic horror tale and its illustrations using AI technologies. This project demonstrates the use of AI, particularly GPT-3 and guided diffusion, to craft a captivating Gothic horror tale with accompanying illustrations. Another paper that conducts the same discussion from the perspective of journalism is a critical analysis by Andrey Miroshnichenko (2018) on the use and impact of artificial intelligence (AI) in the news media. The paper examines how writing algorithms, also known as robo-journalists, can generate natural language texts from structured data or predefined templates, and how they are replacing human journalists in many domains of news production.

Lee's study (2022) examines the concept of creativity in relation to the creative industries, artificial intelligence (AI) and everyday creativity. The author examines how AI tools developed by technology start-ups are being adopted by Hollywood studios to analyse financial, script, and audience data, and to influence their commissioning choices.

Another paper by Chow (2020) begins by providing a brief overview of AI and machine learning, and how they have been applied in various sectors and creative fields. The author concludes by suggesting that we need to rethink creativity in a more holistic and humanistic way, by acknowledging the diverse and complex factors that shape creative processes and outcomes. Another study (Anantrasirichai & Bull, 2021, pp. 590-653) that compares human and AI creativity examines the notion of AI creativity and its

relation to human creativity. They claim that "AI and human creativity are both based on the same underlying principles of information processing, learning, and adaptation".

According to Momot (2022), the film industry is facing imminent changes due to the integration of artificial intelligence (AI) in the filmmaking process. The author suggests that AI scriptwriting will be more useful for generating ideas, suggestions, or feedback for human scriptwriters, rather than producing complete scripts. Similarly, Gümüş and Kocabıyık (2023) argue that AI is not a threat to human creativity, but rather a complementary tool that can enhance and support human writers. The article shows that AI script writers lack the ability to understand the context, meaning, and emotions of the stories that they generate. Nieto (2023) addresses the same question of whether artificial intelligence (AI) can write a screenplay and evaluates three programs that are designed for this purpose: ChatGPT, ScriptBook, and The Safezone. The author argues that AI can be a useful tool for screenwriters, but it cannot replace human creativity and input.

Another study (Infocus Film School, 2023) seeking answers to the same question delves into the inquiry of whether artificial intelligence (AI) can replace human screenwriters in the film industry and provides eight reasons why it cannot. The article evaluates each program based on its strengths and weaknesses and provides examples of their outputs. The article concludes that AI can be a useful tool for screenwriters, but it cannot replace human creativity and input.

A review of the existing studies reveals that the authors agree that AI can play a significant role in screenwriting, assist screenwriters, but not replace them. This research, which seeks to find the answer to the same question, differs from the previous ones by seeking the answer from the participants. Thus, instead of theoretical inferences, this study aims to provide an empirical answer to this question. Hence, this study is unique as the first study in this field that is based on the participants' responses. In this sense, it will help answer many questions in this field, and reveal the current stage of artificial intelligence in screenwriting through practice. Thus, it will identify its current power and capabilities, as well as its weaknesses and shortcomings, and its future potential.

2. Methodology
2.1. Data collection

This research study used screenwriters, an artificial intelligence model (ChatGPT), and participants as data sources. The screenwriters and the artificial intelligence model created the screenplay elements (logline, character introduction, synopsis, treatment, screenplay scene, and dialogue examples), while the participants evaluated these contents and provided feedback. The participants, selected by purposive sampling method, were a group of people who attended a screenplay course.

Various tools were used in the data collection process. The purposive sampling method was used to select a group of people who attended screenplay courses, while standardized open-ended interviews were used to collect the participants' responses and understand their thoughts on the distinction ability between the content created by AI and humans. The analysis of the collected data was conducted by mixed method, both quantitatively and qualitatively. All these processes were carried out based on screenwriting.

First, the screenwriting elements that served the main purpose of the study, which was to distinguish between artificial intelligence and human-produced scenarios, were determined. These elements were logline (story line), synopsis, treatment scene, screenplay scene and dialogue, which constituted all stages of the screenplay.

For the professional screenwriters, a screenwriter who worked for national channels was chosen, without using any parameter, preferring "any Turkish" screenwriter for the research to yield reliable results. This screenwriter was contacted by phone, informed about the research, and consented, and asked to produce content for each of the elements mentioned above by email. The screenwriter's

responses were recorded on a form. Similarly, one of the most advanced artificial intelligence models, ChatGPT 4, was accessed via Bing Chat and asked to write each of these elements in creative mode. This study was conducted in Turkish, which was the native language of the participants. To avoid any bias and not to impair the validity of the results, ChatGPT was given commands as simple as possible, such as "write a very striking and original logline that is your own idea", leaving the creativity to it. The responses of ChatGPT 4 and human screenwriter for each question were merged on a word document under the question title of which one is artificial intelligence, and which one is human screenwriter, thus creating an interview form.

To obtain more accurate results, a question was asked at the beginning of the form about whether the participants had used artificial intelligence before. Also, the question of whether artificial intelligence can replace human screenwriter, which constitutes the subject of the study, was included at the beginning of the form. Then, the productions of artificial intelligence and human screenwriter were complicated and given their final form to the questions. Thus, in each question, logline, character introduction, synopsis, treatment, and screenplay examples produced by human and artificial intelligence were given without revealing who wrote them, and the participants were asked to identify who wrote which one and what they based their identification on. This form consisting of all stages of the screenplay aimed to determine to what extent human and artificial intelligence production screenplay productions can be distinguished from each other.

2.2. Population and sample

The population of the research was everyone who had received screenplay training. The sample was selected from 24 cinema students who had received screenplay training. Purposive sampling method was used in the sample selection. Purposive sampling is a sampling method used in qualitative research. Purposive sampling allows the researcher to select individuals or groups of individuals who have specific criteria that are relevant to the research question or purpose. Purposive sampling enables the researcher to choose the most informative and useful cases based on his/her expertise and knowledge (Tongco, 2007).

The participants of the research were especially those who had received screenplay training and were cinema students. Indeed, it was anticipated that the participants who already knew the stages of the screenplay would focus only on the original source to distinguish between artificial intelligence and human-produced screenplay elements. The fact that the participants had received screenplay training enabled them to know the technical screenplay terms in the form and it was calculated that more consistent results would be obtained in this regard. Gender difference was not considered among the participants. The general age range of the participants was between 20-25.

2.3. Survey method

This study used in-depth interview method, which is one of the qualitative research methods. The structured interview type, which is also known as standard interview and researcher-managed interview (Nor Rashidi, Ara Begum, Mokhtar, & Pereira, 2014), was preferred for this method. Interview is a common method of collecting information from people (Kumar, 2011, p. 137). In-depth interview is a data collection technique frequently used in qualitative research (Uslu & Demir, 2023). This method, which is preferred in qualitative studies, is based on the social interaction and communication between the researcher and the participant. The main goal of in-depth interviews is to reveal the participants' experiences, thoughts and perceptions, to discover the unknowns and to gain new insights (Uslu & Demir, 2023, p. 289). Structured interviews are interviews that the researcher directs to the participants using a pre-prepared set of questions and using the same questions and order (Nor Rashidi, Ara Begum, Mokhtar, & Pereira, 2014; Patton, 1987, p. 112). The list of these questions, which are usually closed-ended and allow little room for interpretation (Demir, 2011, p. 279; Yıldırım & Şimşek 2000, p. 93), is a

written document prepared for person-to-person interviews that take place face-to-face, by phone or by other electronic communication means. The interview program serves as a data collection tool, while the interview refers to the data collection method. One of the biggest advantages of structured interviews, which are especially recommended for studies that require a large number of participants (Yıldırım & Şimşek 2000, p. 96), is that they provide information consistency by making the data comparable. In addition, structured interviews require less interview experience than unstructured interviews (Kumar, 2011, pp. 137-138).

This method can be conducted in two different ways: questionnaire or interview. A questionnaire is a written list of questions that are answered by the participants. In a questionnaire, the participants read, interpret and then write their answers (Kumar, 2011, p. 138). These types of interviews usually offer specific and fixed answer options, are fast and easy. The participants only mark their chosen answers in a box. The purpose of these types of interviews is to ensure that the same questions are asked in the same order in each interview session. Also, the context of the questions is the same for all participants, which facilitates making comparisons with respect to the sample from which data are collected (Nor Rashidi, Ara Begum, Mokhtar, & Pereira, 2014; Bryman, 2012). In this study, as Kumar suggested, group administration form of structured interview type of in-depth interview method was applied. This way it will be revealed whether participants can distinguish between human-produced scenario and artificial intelligence-produced scenario or not.

2.4. Data analysis

This study adopted a mixed methodology in the data analysis process. This approach combines both quantitative and qualitative research methods to provide a more comprehensive and in-depth analysis. The quantitative analysis process involves coding the participants' responses as correct or incorrect and calculating an accuracy rate based on these codings. This process enables obtaining a general success rate and determining which story elements are more difficult or easy to distinguish. The qualitative analysis process involves analyzing the responses to open-ended questions to understand what features or clues the participants based their decisions on. This mixed method approach provides researchers with the opportunity to test theories, conduct deductive and inductive analysis, and address complex phenomena from both subjective and measurable perspectives (Williams, 2007). As a result of this type of analysis, it is aimed to obtain clear findings regarding the evaluation of the differences between artificial intelligence and human-generated scenario elements, which are the subject of the study.

2.5. Interview structure

The interview form, as detailed above, consisting of closed-ended questions and their open-ended interpretations, was given to the participants in written form. The research was conducted using a face-to-face interview method in a school environment. It is important to note that consent declaration forms were obtained from all participants, signed and dated, confirming their voluntary participation in the study as seen in the appendices. Participants were given as much time as they needed to answer the questions, with interviews averaging about one hour in duration. The researcher was present throughout the interview, and during this process, the researcher explained the purpose and importance of the research and was prepared to answer any possible questions. Participants recorded their answers on the interview form, ensuring that the interview was documented. This approach provided a comprehensive record of the participants' responses and allowed for a detailed analysis of their perceptions and experiences. The use of both closed-ended and open-ended questions enabled the collection of both quantitative data (through closed-ended questions) and qualitative data (through open-ended interpretations), providing a rich dataset for analysis.

2.6. Ethics committee permission

Within the framework of the decision taken during the meeting by İstanbul Medipol University Social Sciences Scientific Research Ethics Committee dated 12/13/2023 and numbered 122; the study does not contain any ethical issues.

3. Results and Discussion
3.1. First findings

In this section, the findings from the interviews are presented both quantitatively and qualitatively. Quantitatively, statistical measures such as mean, standard deviation, frequency, and percentage are used to illustrate the accuracy rates of participants in distinguishing between AI-generated and human-generated screenplay elements.

A total of 24 participants were interviewed using the method described above. The first question of the interview was whether they had ever used artificial intelligence before, in order to determine if there was a significant difference in their responses. Of these 24 participants, four had never used artificial intelligence before, while the remaining 20 had. The majority of these had used ChatGPT, suggesting that the use of artificial intelligence, and specifically ChatGPT, is quite widespread.

Similarly, to gauge general opinion, the second question asked was the research question of this paper: Can AI-generated screenplays be distinguished from human-generated ones? Looking at the responses, two of the four participants who had never used artificial intelligence before thought they could be distinguished, while the other two thought they could not be distinguished. One participant thought they could be distinguished before seeing the interview question form, but decided they could not be distinguished after having difficulty distinguishing between the AI-generated options.

Of the 20 participants who had previously used artificial intelligence, 15 believed that AI-generated screenplays could be distinguished from human-generated ones, while four argued they could not be distinguished. One participant did not express an opinion. The overwhelming majority held the view that AI-generated screenplays could be distinguished from human-generated ones.

Among those who expressed the view that AI-generated screenplays could be distinguished from human-generated ones, one participant qualified this by stating that this would change in the future. Another participant who expressed the view that they could be distinguished stated that AI put forth clichéd, stereotypical and popular ideas. Yet another participant who expressed the view that they could be distinguished argued conversely that AI could be distinguished from a human screenwriter because it also has a style like human screenwriters.

The third question directed at participants was: Can AI-generated screenplays replace human-generated ones in the future? Looking at the responses, all four participants who had never used artificial intelligence before stated that AI could replace human screenwriters. One of these participants stipulated that AI would need to be able to mimic emotion in order to replace human screenwriters. The other participants also stated that AI could replace human screenwriters but would need further development. Of the 20 participants who had previously used artificial intelligence, 14 argued that AI could replace human screenwriters in the future, while five argued it could not.

Among those who said AI could not replace human screenwriters, one participant stated that human control and guidance were essential in screenplay production by AI. Other participants who expressed the view that AI could replace human screenwriters put forth this idea on condition of further development of AI. Two participants who expressed the view that it could replace them advocated for both AI-generated and human-generated screenplays coexisting. In general terms, the common opinion emerged that under current conditions AI-generated screenplays can be distinguished from human-

generated ones but with future developments in AI, it may replace human screenwriters. According to statistical interpretation of our first three interview findings:

The first question of the interview was whether they had ever used artificial intelligence before. This question was used to determine if there was a significant difference in their responses based on their prior experience with artificial intelligence. The results showed that 83.3% (20 out of 24) of the participants had used artificial intelligence before, while only 16.7% (4 out of 24) had not. This suggests that the use of artificial intelligence is quite common among the participants.

When asked whether AI-generated screenplays can be distinguished from human-generated ones, 50% of the participants who had never used AI before believed they could be distinguished, while the other 50% thought they could not. The results showed that 75% (15 out of 20) of the participants who had used artificial intelligence before believed that they could distinguish AI-generated screenplays from human-generated ones, while only 20% (4 out of 20) argued that they could not. One participant did not express an opinion. This indicates that most of the participants who had used artificial intelligence before were confident in their ability to differentiate AI-generated screenplays from human-generated ones.

Regarding the question of whether AI-generated screenplays can replace human-generated ones in the future, all participants who had never used AI before believed that AI could replace human screenwriters. For the other participants the results showed that 100% (4 out of 4) of the participants who had never used artificial intelligence before stated that AI could replace human screenwriters, while 70% (14 out of 20) of the participants who had used artificial intelligence before agreed with this statement. However, 25% (5 out of 20) of the participants who had used artificial intelligence before disagreed with this statement, arguing that AI could not replace human screenwriters. This suggests that there is a difference in the perception of AI's potential between the participants who had never used artificial intelligence before and those who had. The participants who had never used artificial intelligence before were more optimistic about AI's ability to replace human screenwriters, while the participants who had used artificial intelligence before were more cautious and critical. Some of the participants also mentioned that AI would need to be able to mimic emotion and creativity in order to replace human screenwriters.

In summary, the responses to the question of whether AI-generated screenplays can be distinguished from human-generated ones suggest that prior experience with AI does not significantly influence the ability to distinguish between these two types of screenplays. This could imply that the distinguishing features of AI-generated screenplays are noticeable regardless of one's familiarity with AI. A majority of participants, both those who had and had not previously used AI, believed that AI could potentially replace human screenwriters in the future. This indicates a general optimism or openness towards the integration of AI in creative fields such as screenwriting. Despite the general trends, there is a diversity of opinions among participants. Some participants qualified their responses by stating that the capabilities of AI would need to improve for it to replace human screenwriters, or that they expected the distinction between AI-generated and human-generated screenplays to become less clear in the future. This highlights the complexity of the issue and suggests that while AI has potential in this field, its role and impact are still subjects of debate.

The results emerging from these responses align in some ways with the claims made in previous studies mentioned in the literature review, as those researchers also argue that works produced by AI can be distinguished from human-produced screenplays. However, unlike this study, a majority of those researchers argue that even in the future, screenplay production by AI will only be possible with human direction or with human creation at its core. The debate followed through screenplay elements below will be able to reveal more accurate results. Indeed, both the responses listed above from participants and the validity of previous researchers' claims can be tested practically.

3.2. Logline: human screenwriter versus AI: who is the creator?

The participants were shown four loglines, two of which were AI-generated and two of which were human-generated. The participants were given the following loglines and were asked to determine which ones were generated by artificial intelligence and which ones were created by a human screenwriter. AI-generated Logline 1: "A spaceship lands on a distant planet. The planet's inhabitants welcome the crew of the ship. However, when they discover the secret of the planet, they have to fight to survive: The planet is actually the body of a giant monster." AI-generated Logline 2: "A group of young people gather in a house to play a mysterious game. However, the game confronts them with a horrific reality: according to the rules of the game, they have to kill each other before they can leave the house." Human-generated Logline 1: "A young woman who is about to get married finds out that her fiancé is her past abuser from the flash drive she got in her hand." Human-generated Logline 2: "Zeinyx, who is about to save the world, finds himself as Zahit lying in the hospital." The results were as follows:

Table 1

Participant Identification of AI-Generated and Human-Generated Loglines

Logline Type	Participant Experience with AI	Correctly Identified as AI-Generated	Incorrectly Identified as Human-Generated	Correctly Identified as Human-Generated	Incorrectly Identified as AI-Generated
AI-Generated Logline 1	Used AI	75% (15/20)	25% (5/20)	-	-
	Never Used AI	75% (3/4)	25% (1/4)	-	-
	Total	**91.7% (22/24)**	**8.3% (2/24)**	-	-
AI-Generated Logline 2	Used AI	40% (8/20)	60% (12/20)	-	-
	Never Used AI	0% (0/4)	100% (4/4)	-	-
	Total	**33.3% (8/24)**	**66.7% (16/24)**	-	-
Human-Generated Logline 1	Used AI	-	-	95% (19/20)	5% (1/20)
	Never Used AI	-	-	75% (3/4)	25% (1/4)
	Total	-	-	**91.7% (22/24)**	**8.3% (2/24)**
Human-Generated Logline 2	Used AI	-	-	55% (11/20)	45% (9/20)
	Never Used AI	-	-	25% (1/4)	75% (3/4)
	Total	-	-	**50% (12/24)**	**50% (12/24)**

The data shows that the participants' ability to distinguish human-generated loglines and AI-generated loglines varied depending on the quality and complexity of the loglines. Some loglines were easier to identify than others, and some loglines were more confusing than others.

For the first logline, which was AI-generated, the majority of the participants correctly guessed that it was AI-generated, regardless of their prior experience with artificial intelligence. This suggests that the first logline was relatively easy to identify as AI-generated, perhaps because it had some features that were typical of AI-generated texts, such as grammatical errors, logical inconsistencies, or lack of originality. For the second logline, which was also AI-generated, the majority of the participants incorrectly guessed that it was human-generated, regardless of their prior experience with artificial intelligence. This suggests that the second logline was relatively difficult to identify as AI-generated, perhaps because it had some features that were typical of human-generated texts, such as coherence, creativity, or emotion. For the first logline, which was human-generated, the majority of the participants

correctly guessed that it was human-generated, regardless of their prior experience with artificial intelligence. This suggests that the third logline was relatively easy to identify as human-generated, perhaps because it had some features that were typical of human-generated texts, such as plot, character, or genre. For the second logline, which was also human-generated, the participants were evenly split between guessing that it was human-generated and guessing that it was AI-generated, regardless of their prior experience with artificial intelligence. This suggests that the fourth logline was relatively confusing to identify as human-generated or AI-generated, perhaps because it had some features that were ambiguous or mixed, such as style, tone, or theme.

Statistically, this means that the participants correctly guessed 54.2% of the loglines, regardless of whether they were AI-generated or human-generated. This is slightly better than random guessing, which would have an accuracy of 50%. However, the overall accuracy does not tell us the difference between the accuracy of guessing AI-generated loglines and the accuracy of guessing human-generated loglines. But the participants correctly guessed 62.5% of the AI-generated loglines and 70.8% of the human-generated loglines. This shows that the participants were more accurate in guessing human-generated loglines than AI-generated loglines. This could imply that the human-generated loglines had slightly more distinctive features that made them easier to identify, or that the AI-generated loglines had more deceptive features that made them harder to identify. Based on the numerical and statistical analysis, I can say that the rate of distinguishing AI-generated screenplays from human-generated ones is 54.2%, which is slightly better than random guessing. However, I cannot say that this rate is significantly different for AI-generated loglines and human-generated loglines, as the chi-square test of independence did not show a significant relationship between the type of logline and the type of guess. Therefore, the answer to the question of whether human-generated loglines and AI-generated loglines can be distinguished from each other is not clear-cut, and it may depend on other factors, such as the quality and complexity of the loglines, the prior experience and knowledge of the participants, and the context and purpose of the loglines.

3.3. Logline: human screenwriter versus AI: which one is better?

It was requested to rank the above loglines from the most liked one (starting from 1) to the least liked one (up to 4) in order to determine whether the one produced by artificial intelligence or the one created by a human screenwriter received more favorable responses. The preference levels for each logline were as follows:

Table 2

Preference Ranking of AI-Generated and Human-Generated Loglines

Logline Description	Score 1	Score 2	Score 3	Score 4	Average Score	Preference Level
AI-Generated Logline 1	8.3% (2/24)	45.8% (11/24)	37.5% (9/24)	8.3% (2/24)	2.58	Moderate
AI-Generated Logline 2	12.5% (3/24)	45.8% (11/24)	20.8% (5/24)	8.3% (2/24)	2.38	Low
Human-Generated Logline 1	16.7% (4/24)	33.3% (8/24)	29.2% (7/24)	16.7% (4/24)	2.5	Moderate
Human-Generated Logline 2	29.2% (7/24)	20.8% (5/24)	16.7% (4/24)	25% (6/24)	2.54	Moderate

This means that the average score for human-generated loglines was slightly higher than the average score for AI-generated loglines, by 0.04 points. This suggests that the participants preferred human-generated loglines over AI-generated loglines, but not by a large margin. This could imply that the human-generated loglines had more appealing features that made them more liked, or that the AI-generated loglines had more unappealing features that made them less liked. This small margin could

also imply that the difference in the average scores for AI-generated loglines and human-generated loglines is due to chance or other factors, and not due to the type of logline itself.

When we look at all these results, we cannot find a clear answer to the question of whether the logline was produced by artificial intelligence or human, or to the question of which logline was more liked. This generally opens the way for an inference that artificial intelligence and human-generated loglines can replace each other.

3.4. Synopsis: human screenwriter versus AI: who is the creator?

The participants were shown four synopses, two of which were AI-generated and two of which were human-generated. They were asked to guess which ones belonged to whom. The results were as follows:

Table 3

Participant Identification of Human-Generated and AI-Generated Synopses

Synopsis Type	Participant Experience with AI	Correctly Identified as AI-Generated	Incorrectly Identified as Human-Generated	Correctly Identified as Human-Generated	Incorrectly Identified as AI-Generated
Human-Generated Synopsis 1	Used AI	-	-	55% (11/20)	45% (9/20)
	Never Used AI	-	-	25% (1/4)	75% (3/4)
	Total	-	-	**54.2% (13/24)**	**45.8% (11/24)**
Human-Generated Synopsis 2	Used AI	-	-	70% (14/20)	30% (6/20)
	Never Used AI	-	-	0% (0/4)	100% (4/4)
	Total	-	-	**58.3% (14/24)**	**41.7% (10/24)**
AI-Generated Synopsis 1	Used AI	45% (9/20)	55% (11/20)	-	-
	Never Used AI	75% (3/4)	25% (1/4)	-	-
	Total	**50% (12/24)**	**50% (12/24)**	-	-
AI-Generated Synopsis 2	Used AI	15% (3/20)	85% (17/20)	-	-
	Never Used AI	0% (0/4)	100% (4/4)	-	-
	Total	**12.5% (3/24)**	**87.5% (21/24)**	-	-

This means that the participants correctly guessed 31.3% of the AI-generated synopses and 56.3% of the human-generated synopses. This shows that the participants were more accurate in guessing human-generated synopses than AI-generated synopses. This could imply that the human-generated synopses had more distinctive features that made them easier to identify, or that the AI-generated synopses had more deceptive features that made them harder to identify.

This means that the participants correctly guessed 44.8% of the synopses, regardless of whether they were AI-generated or human-generated. This is worse than random guessing, which would have an accuracy of 50%. This also means that the participants had a low ability to distinguish human-generated synopses and AI-generated synopses. They were more likely to make a wrong guess than a right guess, regardless of their prior experience with artificial intelligence. This could imply that the AI-generated synopses were very similar to the human-generated ones, or that the human-generated synopses were very different from each other. It could also imply that the participants were not familiar with the features and criteria that could help them identify the type of synopsis. Therefore, this result indicates

that the current version of artificial intelligence (GPT 4) can produce synopses that are hard to differentiate from human-generated ones.

3.5. Synopsis: human screenwriter versus AI: which one is better?

The participants were asked to rank the four synopses from the most liked to the least liked, using numbers from 1 to 4. The preference levels for each synopsis were as follows:

Table 4

Preference Ranking of Human-Generated and AI-Generated Synopses

Synopsis Description	Score 1	Score 2	Score 3	Score 4	Average Score	Preference Level
Human-Generated Synopsis 1	12.5% (3/24)	50% (12/24)	12.5% (3/24)	25% (6/24)	2.5	Moderate
Human-Generated Synopsis 2	4.2% (1/24)	41.7% (10/24)	33.3% (8/24)	12.5% (3/24)	2.63	Moderate
AI-Generated Synopsis 1	20.8% (5/24)	25% (6/24)	33.3% (8/24)	16.7% (4/24)	2.5	Moderate
AI-Generated Synopsis 2	12.5% (3/24)	25% (6/24)	41.7% (10/24)	20.8% (5/24)	2.71	Moderate

This means that the average score for AI-generated synopses was slightly higher than the average score for human-generated synopses, by 0.04 points. This suggests that the participants preferred AI-generated synopses over human-generated synopses, but not by a large margin. This could imply that the AI-generated synopses had more appealing features that made them more liked, or that the human-generated synopses had more unappealing features that made them less liked.

3.6. Character analysis: Human screenwriter versus AI: who is the creator?

The participants were shown four character descriptions, two of which were AI-generated and two of which were human-generated. They were asked to guess which ones belonged to whom. The results were as follows:

Table 5

Participant Identification of Human-Generated and AI-Generated Character Descriptions

Character Description	Participant Experience with AI	Correctly Identified as AI-Generated	Incorrectly Identified as Human-Generated	Correctly Identified as Human-Generated	Incorrectly Identified as AI-Generated
Human-Generated Description 1	Used AI	45% (9/20)	55% (11/20)	-	-
	Never Used AI	50% (2/4)	50% (2/4)	-	-
	Total	**54.2% (13/24)**	**45.8% (11/24)**	-	-
Human-Generated Description 2	Used AI	25% (5/20)	75% (15/20)	-	-
	Never Used AI	75% (3/4)	25% (1/4)	-	-
	Total	**62.5% (15/24)**	**37.5% (9/24)**	-	-
AI-Generated Description 1	Used AI	35% (7/20)	65% (13/20)	-	-
	Never Used AI	0% (0/4)	100% (4/4)	-	-
	Total	**29.2% (7/24)**	**70.8% (17/24)**	-	-

Character Description	Participant Experience with AI	Correctly Identified as AI-Generated	Incorrectly Identified as Human-Generated	Correctly Identified as Human-Generated	Incorrectly Identified as AI-Generated
AI-Generated Description 2	Used AI	40% (8/20)	60% (12/20)	-	-
	Never Used AI	75% (3/4)	25% (1/4)	-	-
	Total	**45.8% (11/24)**	**54.2% (13/24)**	-	-

This means that the participants correctly guessed 47.9% of the character descriptions, regardless of whether they were AI-generated or human-generated. This is worse than random guessing, which would have a correct guess rate of 50%. This indicates that the participants had a low ability to distinguish human-generated character descriptions and AI-generated character descriptions. Based on the data, I can conclude that human-generated character descriptions and AI-generated character descriptions cannot be distinguished from each other.

3.7. Character analysis: human screenwriter versus AI: which is better?

The participants were asked to rank the four character descriptions from the most liked to the least liked, using numbers from 1 to 4. The preference levels for each character description were as follows:

Table 6

Preference Ranking of Human-Generated and AI-Generated Character Descriptions

Character Description	Score 1	Score 2	Score 3	Score 4	Average Score	Preference Level
Human-Generated Description 1	12.5% (3/24)	29.2% (7/24)	33.3% (8/24)	20.8% (5/24)	2.67	Moderate
Human-Generated Description 2	16.7% (4/24)	37.5% (9/24)	20.8% (5/24)	12.5% (3/24)	2.42	Low
AI-Generated Description 1	4.2% (1/24)	33.3% (8/24)	25% (6/24)	33.3% (8/24)	2.92	High
AI-Generated Description 2	29.2% (7/24)	25% (6/24)	25% (6/24)	12.5% (3/24)	2.29	Low

To answer the question of whether human-generated character descriptions or AI-generated character descriptions were more liked, I have compared the average scores for each type of character description.

This means that the average score for AI-generated character descriptions was slightly higher than the average score for human-generated character descriptions, by 0.06 points. This suggests that the participants preferred AI-generated character descriptions over human-generated character descriptions, but not by a large margin.

3.8. Treatment scene: Human screenwriter versus AI: who is the creator?

The participants were shown four treatment scenes, two of which were AI-generated and two of which were human-generated. They were asked to guess which ones belonged to whom. The results were as follows:

Table 7

Participant Identification of Human-Generated and AI-Generated Treatment Scenes

Treatment Scene	Participant Experience with AI	Correctly Identified as AI-Generated	Incorrectly Identified as Human-Generated	Correctly Identified as Human-Generated	Incorrectly Identified as AI-Generated
Human-Generated Scene 1	Used AI	25% (5/20)	75% (15/20)	-	-
	Never Used AI	25% (1/4)	75% (3/4)	-	-
	Total	**75% (18/24)**	**25% (6/24)**	-	-
Human-Generated Scene 2	Used AI	25% (5/20)	75% (15/20)	-	-
	Never Used AI	25% (1/4)	75% (3/4)	-	-
	Total	**75% (18/24)**	**25% (6/24)**	-	-
AI-Generated Scene 1	Used AI	55% (11/20)	45% (9/20)	-	-
	Never Used AI	50% (2/4)	50% (2/4)	-	-
	Total	**54.2% (13/24)**	**45.8% (11/24)**	-	-
AI-Generated Scene 2	Used AI	70% (14/20)	30% (6/20)	-	-
	Never Used AI	50% (2/4)	50% (2/4)	-	-
	Total	**66.7% (16/24)**	**33.3% (8/24)**	-	-

This means that the participants correctly guessed 68.8% of the treatment scenes, regardless of whether they were AI-generated or human-generated. This is better than random guessing, which would have a correct guess rate of 50%. This indicates that the participants had a high ability to distinguish human-generated treatment scenes and AI-generated treatment scenes. Treatment scenes can be distinguished more easily than the loglines, synopses, and character descriptions.

3.9. Treatment scene: human screenwriter versus AI: which one is better?

The participants were asked to rank the four treatment scenes from the most liked to the least liked, using numbers from 1 to 4. The preference levels for each treatment scene were as follows:

Table 8

Preference Ranking of Human-Generated and AI-Generated Treatment Scenes

Treatment Scene	Score 1	Score 2	Score 3	Score 4	Average Score	Preference Level
Human-Generated Scene 1	29.2% (7/24)	25% (6/24)	25% (6/24)	12.5% (3/24)	2.29	Low
Human-Generated Scene 2	16.7% (4/24)	33.3% (8/24)	33.3% (8/24)	16.7% (4/24)	2.5	Moderate
AI-Generated Scene 1	12.5% (3/24)	33.3% (8/24)	33.3% (8/24)	20.8% (5/24)	2.63	Moderate
AI-Generated Scene 2	20.8% (5/24)	29.2% (7/24)	33.3% (8/24)	8.3% (2/24)	2.38	Low

To answer the question of whether human-generated treatment scenes or AI-generated treatment scenes were more liked, I have compared the average scores for each type of treatment scene. This means that the average score for AI-generated treatment scenes was slightly higher than the average score for human-generated treatment scenes, by 0.11 points. This suggests that the participants

preferred AI-generated treatment scenes over human-generated treatment scenes, but not by a large margin. This means that the participants did not have a strong preference for either type of treatment scene, and that their preference levels were similar for both types.

3.10. Script scene: human screenwriter versus AI: who is the creator?

The participants were shown four screenplay scenes, two of which were AI-generated and two of which were human-generated. They were asked to guess which ones belonged to whom. The results were as follows:

Table 9

Participant Identification of Human-Generated and AI-Generated Screenplay Scenes

Screenplay Scene	Participant Experience with AI	Correctly Identified as AI-Generated	Incorrectly Identified as Human-Generated	Correctly Identified as Human-Generated	Incorrectly Identified as AI-Generated
Human-Generated Scene 1	Used AI	30% (6/20)	70% (14/20)	-	-
	Never Used AI	50% (2/4)	50% (2/4)	-	-
	Total	**66.7% (16/24)**	**33.3% (8/24)**	-	-
Human-Generated Scene 2	Used AI	40% (8/20)	60% (12/20)	-	-
	Never Used AI	75% (3/4)	25% (1/4)	-	-
	Total	**50% (12/24)**	**50% (12/24)**	-	-
AI-Generated Scene 1	Used AI	30% (6/20)	70% (14/20)	-	-
	Never Used AI	0% (0/4)	100% (4/4)	-	-
	Total	**25% (6/24)**	**75% (18/24)**	-	-
AI-Generated Scene 2	Used AI	60% (12/20)	40% (8/20)	-	-
	Never Used AI	25% (1/4)	75% (3/4)	-	-
	Total	**54.2% (13/24)**	**45.8% (11/24)**	-	-

The participants correctly guessed 47.9% of the screenplay scenes, regardless of whether they were AI-generated or human-generated. This is worse than random guessing, which would have a correct guess rate of 50%. This indicates that the participants had a low ability to distinguish human-generated screenplay scenes and AI-generated screenplay scenes. Based on the data, I can conclude that human-generated screenplay scenes and AI-generated screenplay scenes cannot be distinguished from each other easily and consistently.

3.11. Script scene: human screenwriter versus AI: which one is better?

The participants were asked to rank the four screenplay scenes from the most liked to the least liked, using numbers from 1 to 4. The preference levels for each screenplay scene were as follows:

Table 10

Preference Ranking of Human-Generated and AI-Generated Screenplay Scenes

Screenplay Scene	Score 1	Score 2	Score 3	Score 4	Average Score	Preference Level
Human-Generated Scene 1	16.7% (4/24)	33.3% (8/24)	25% (6/24)	25% (6/24)	2.58	Moderate
Human-Generated Scene 2	20.8% (5/24)	33.3% (8/24)	25% (6/24)	25% (6/24)	2.5	Moderate
AI-Generated Scene 1	4.2% (1/24)	41.7% (10/24)	29.2% (7/24)	20.8% (5/24)	2.71	Moderate
AI-Generated Scene 2	16.7% (4/24)	12.5% (3/24)	41.7% (10/24)	12.5% (3/24)	2.67	Moderate

To answer the question of whether human-generated screenplay scenes or AI-generated screenplay scenes were more liked, I have compared the average scores for each type of screenplay scene. According to results, the average score for AI-generated screenplay scenes was slightly higher than the average score for human-generated screenplay scenes, by 0.15 points. This suggests that the participants preferred AI-generated screenplay scenes over human-generated screenplay scenes, but not by a large margin.

3.12. Dialog writing

Before delving into the tables, the examination of dialog writing illustrates how distinguishing between artificial intelligence (AI) and human-generated content can be quite intricate. In this section, we focus on two tables containing accuracy rates based on participants' experience and lack thereof with AI. These data underscore the complexity of discerning between AI-generated and human-generated content, influenced by participants' backgrounds and prior exposure to AI.

Table 11

Participant Accuracy in Identifying AI-Generated and Human-Generated Dialogues

Participant Group	Experience with AI	Total Guesses	Correct Guesses	Accuracy (%)
Participants with AI Experience	Used AI	150	70	46.67
Participants without AI Experience	Never Used AI	24	13	54.17
Overall Average Accuracy	-	176	83	47.22

Table 12

Dialogue Identification Accuracy by Dialogue Type

Dialogue Type	Total Guesses	Correct Guesses	Accuracy (%)
AI-Generated Dialogues	114	37	32.41
Human-Generated Dialogues	69	46	66.67

The dialogues produced by artificial intelligence were more difficult to identify than the ones produced by human scriptwriters. The average accuracy for the artificial intelligence dialogues was 32.41%, while the average accuracy for the human scriptwriter dialogues was 66.67%. These comparisons suggest that the artificial intelligence dialogues were more challenging to identify than the human scriptwriter dialogues. Additionally, participants who had not previously used artificial intelligence were slightly more accurate in their identifications than those who had. This implies that artificial intelligence was capable of producing dialogues similar to those of human scriptwriters but not identical. Moreover, previous experience with artificial intelligence may have influenced participants' judgments, leading them to be more inclined to guess human production when uncertain.

Therefore, based on these data, we cannot say that artificial intelligence and human production can be easily distinguished. However, we can also not say that they are indistinguishable, because there were some differences in the accuracy rates. A possible conclusion is that artificial intelligence and human production can be partially distinguished, depending on the quality of the dialogues and the background of the participants.

4. Limitations and Challenges

This study had its some difficulties as it did not have a similar study to follow and had to create its own genre by trial and error. First of all, there were three different sources for collecting the data: human scriptwriter, artificial intelligence, and participants. It was inevitable that each of these sources had their own problems. Primarily, according to what criteria would the human scriptwriter write the script elements? We could not draw a boundary by determining a topic, as it would harm the objectivity of the study. We left the human scriptwriter completely free for this. Likewise, we did not give a topic to artificial intelligence. However, since artificial intelligence works more on specific instructions, we actually limited its creativity to some extent. Still, both artificial intelligence and human scriptwriter produced the script elements without any intervention. In fact, we knew that if we gave more specific instructions to artificial intelligence, it would produce much more successful results. However, in that case, it would be artificial intelligence-human co-production. This would make it impossible to compare human scriptwriter and artificial intelligence. For this reason, the data were obtained without giving any special requests or commands to both human and artificial intelligence. While doing this, in order not to get a random result, all script elements were included: logline, treatment, synopsis, character, dialogue. Although it was difficult and circuitous, this path proved the correctness of the chosen method, as the results obtained were close to each other and confirmed each other. Similar problems were experienced with the participants. Some of the participants (4 participants) had not used artificial intelligence before, while the majority (20 participants) had used artificial intelligence. We also had to consider the participants' prior experience with artificial intelligence. We wondered if there would be a difference in the detection of artificial intelligence-generated script elements between those who had used artificial intelligence and those who had not. We also included this topic, which could be an article topic, and presented the results. In summary, despite the difficulties experienced in all three data sources, the results obtained were consistent with each other.

5. Conclusion

The study conducted an in-depth analysis comparing human-generated and AI-generated screenwriting in terms of identification accuracy, preference rankings, and overall effectiveness. The findings contribute to the ongoing debate on the capabilities of artificial intelligence in creative domains, particularly in screenplay writing.

In summary, looking at the numerical data, human-generated script elements and AI-generated script elements cannot be easily distinguished. The average accuracy rate for all the elements is 52.7% (253 out of 480 correct guesses), which is slightly higher than random guessing (50%). This means that the participants were not very confident or consistent in their judgments, and that the AI was able to produce script elements that were similar to the human-generated ones, but not identical.

The most distinguishable elements are the treatment scenes, the loglines, and the synopses, in that order. The treatment scenes had the highest accuracy rate of 68.8% (165 out of 240 correct guesses), which means that the participants were able to identify the AI-generated scenes more often than the human-generated ones. The loglines had the second highest accuracy rate of 54.2% (65 out of 120 correct guesses), which means that the participants were slightly better at identifying the human-generated loglines than the AI-generated ones. The synopses had the third highest accuracy rate of 44.8% (54 out of 120 correct guesses), which means that the participants were slightly worse at

identifying the human-generated synopses than the AI-generated ones. The character descriptions and the screenplay scenes had the lowest accuracy rates of 47.9% (58 out of 120 correct guesses) and 47.2% (57 out of 120 correct guesses), respectively, which means that the participants were almost equally likely to guess human or AI production for these elements.

These numbers suggest that the AI was able to produce script elements that were close to the human-generated ones, but not exactly the same. The AI may have used some patterns or formulas that were common in the script elements, but also made some mistakes or inconsistencies that the human scriptwriter would not. The participants may have used some criteria or intuition to judge the script elements, but also had some biases or uncertainties that affected their judgments. The results also suggest that some script elements are more difficult to generate or identify than others, depending on the complexity, creativity, and specificity of the element.

Looking at the preference data, human-generated script elements and AI-generated script elements have not been able to establish superiority over each other. The average score for all the script elements was 3.51 out of 5, which means that the participants had a moderate level of liking for both types of script elements. The highest score was for the human-generated loglines, which had an average score of 3.67 out of 5. The lowest score was for the AI-generated character descriptions, which had an average score of 3.29 out of 5.

AI has been most successful in writing the screenplay scenes, the treatment scenes, and the synopses, in that order. The screenplay scenes had the highest score for the AI-generated script elements, with an average score of 3.63 out of 5, which means that the participants liked the AI-generated screenplay scenes slightly more than the human-generated ones, which had an average score of 3.48 out of 5. The treatment scenes had the second highest score for the AI-generated script elements, with an average score of 3.54 out of 5, which means that the participants liked the AI-generated treatment scenes slightly more than the human-generated ones, which had an average score of 3.43 out of 5. The synopses had the third highest score for the AI-generated script elements, with an average score of 3.46 out of 5, which means that the participants liked the AI-generated synopses slightly more than the human-generated ones, which had an average score of 3.42 out of 5.

Based on these data, we can say that there is a slight difference between the participants who had previously used artificial intelligence and those who had not, in terms of identifying the AI-generated script elements. The participants who had previously used artificial intelligence had a lower accuracy rate than those who had not, for all the script elements except for the treatment scenes. This means that the participants who had previously used artificial intelligence were more likely to guess human production when they encountered AI-generated script elements, and vice versa. This may indicate that the participants who had previously used artificial intelligence had some biases or expectations that influenced their judgments, and that they were less sensitive or attentive to the differences between the AI-generated and human-generated script elements. On the other hand, the participants who had not previously used artificial intelligence had a higher accuracy rate than those who had, for all the script elements except for the treatment scenes. This means that the participants who had not previously used artificial intelligence were more likely to guess AI production when they encountered AI-generated script elements, and vice versa. This may indicate that the participants who had not previously used artificial intelligence had some curiosity or intuition that guided their judgments, and that they were more aware or critical of the differences between the AI-generated and human-generated script elements. However, the difference between the two groups of participants was not very large, and none of them had a high accuracy rate. This means that the previous experience with artificial intelligence was not a decisive factor in identifying the AI-generated script elements, and that there were other factors that affected the participants' judgments, such as the quality, complexity, and creativity of the script elements.

Based on these elements, we can say that AI can replace human scriptwriter in terms of preference now or in the near future. The preference levels for the AI-generated and human-generated script elements were very close, and none of them had a high level of liking. This means that the AI was able to produce script elements that were acceptable, but not outstanding. The AI may have used some techniques or features that were appealing, but also made some errors or shortcomings that the human scriptwriter would not. The participants may have had some expectations or criteria to evaluate the script elements, but also had some variations or preferences that affected their ratings. The results also suggest that some script elements are more difficult to write or evaluate than others, depending on the complexity, creativity, and specificity of the element. For example, the screenplay scenes may have required more personality, emotion, and dialogue than the other elements, which made them more likable. The loglines may have required more originality, conciseness, and appeal than the other elements, which made them more difficult to write. The character descriptions may have required more consistency, clarity, and structure than the other elements, which made them less likable.

Looking at all these data, this study reveals the potential of artificial intelligence and shows that it has reached a level where it can compete with human scriptwriter. In this sense, the effort of American scriptwriters to impose restrictions on artificial intelligence becomes very meaningful. Moreover, it is also seen that the average opinion in the literature, that artificial intelligence needs human collaboration, is not true. On the contrary, it is the human who needs artificial intelligence collaboration. Considering that artificial intelligence is still at the beginning of its journey, perhaps it is time to treat it not as a supporting actor, but as a deferred soloist. Future research on this topic should be more practical than theoretical, as it will produce more accurate results. Indeed, here we have examined the artificial intelligence production in terms of ChatGPT 4, which is at least a part of, and potentially the future of, scriptwriting. To conclude, this study demonstrates that artificial intelligence is not only a powerful tool, but also a creative partner for human scriptwriter. Artificial intelligence can produce script elements that are comparable to, and sometimes even preferred over, human-generated ones. Artificial intelligence can also challenge and inspire human scriptwriter to improve their skills and styles. Whether artificial intelligence is a threat or not is arguable, but it is the future inevitable.

References

Anantrasirichai, N., & Bull, D. (2021). Artificial intelligence in the creative industries: A review. *Artificial Intelligence Review, 55*(4), 589-656. https://doi.org/10.1007/s10462-021-10039-7

Anguiano, D., & Beckett, L. (2023, October 1). How Hollywood writers triumphed over AI – and why it matters. *The Guardian*.

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.

Chow, P.-S. (2020). Ghost in the (Hollywood) machine: Emergent applications of artificial intelligence in the film industry. *NECSUS_European Journal of Media Studies, 9*(1), 193–214. https://doi.org/10.25969/mediarep/14307

Demir, O. Ö. (2011). Nitel araştırma yöntemleri. In K. Böke (Ed.), *Sosyal bilimlerde araştırma yöntemleri* (pp. 275-306). Alfa Yayınları.

Gümüş, Y. E., & Kocabıyık, Y. (2023). Yapay zekâ teknolojisi yaratıcı yazarlığa karşı: dijital çağda yaratıcı yazarlığın dönüşümü. *The Turkish Online Journal of Design, Art and Communication - TOJDAC, 13*(4), 1178-1191. https://doi.org/10.7456/tojdac.1323219

InFocus Film School. (2023, 8 Ağustos). Will AI replace screenwriters?: 8 reasons AI can't write good scripts. Retrieved from https://infocusfilmschool.com/will-ai-replace-screenwriters/

Kavitha, L. (2023). Copyright challenges in the artificial intelligence revolution: Transforming the film industry from script to screen. *Trinity Law Review, 4*(1), 1-20. https://doi.org/10.48165/tlr.2024.4.1.1

Kumar, R. (2011). *Research methodology: A step-by-step guide for beginners*. Sage Publications.

Kurt, D. E. (2018). Artistic creativity in artificial intelligence (Master's thesis, Radboud University).

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Viking Penguin.

Lee, H.-K. (2022). Rethinking creativity: Creative industries, AI and everyday creativity. *Media, Culture & Society*, 44 (3), 601-612. https://doi.org/10.1177/01634437221077009

Li, Y. (2022). Research on the application of artificial intelligence in the film industry. *SHS Web of Conferences, 144*, 03002. https://doi.org/10.1051/shsconf/202214403002

Marr, B. (2023, May 19). A short history of ChatGPT: How we got to where we are today. *Forbes*. Retrieved from https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/?sh=613f67d7674f

McCarthy, J. (2007, November 12). What is artificial intelligence? Computer Science Department, Stanford University.

Mijwel, M. M. (2015). History of artificial intelligence. *Research Gate* 3, 1-8. https://doi.org/10.13140/RG.2.2.16418.15046

Miroshnichenko, A. (2018). AI to bypass creativity. Will robots replace journalists? (The answer is "Yes"). *Information, 9*(7), 163-183. https://doi.org/10.3390/info9070183

Mirowski, P. W., Mathewson, K. W., Pittman, J., & Evans, R. (2023, April, 23-28). Co-writing screenplays and theatre scripts with language models: Evaluation by industry professionals. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), Hamburg, Germany*. ACM, New York, NY, USA. https://doi.org/10.1145/3544548.3581225

Momot, I. (2022). Artificial intelligence in filmmaking process: Future scenarios [Bachelor's thesis, JAMK University of Applied Sciences].

Nieto, J. (2023, April 20). Can AI write a screenplay? A review of 3 programs. *AMT Lab*. https://amt-lab.org/reviews/2023/4/can-ai-screenwrite-a-review-of-3-programs

Nor Rashidi, M., Ara Begum, R., Mokhtar, M., & Pereira, J. J. (2014). The conduct of structured interviews as research implementation method. *Journal of Advanced Research Design, 1*(1). 18-27.

Patton, Q. M. (1987). *How to use qualitative methods in evaluation*. Sage Publications.

Shah, S. (2023, May 4). Why the WGA is striking for limits on the use of AI. *Time*. https://time.com/6277158/writers-strike-ai-wga-screenwriting/

Song, Y. (2022, June 1). Analysis on whether artificial intelligence can replace human screenwriters. In *Proceedings of the 2022 8th International Conference on Humanities and Social Science Research (ICHSSR 2022) (Advances in Social Science, Education and Humanities Research)*. Atlantis Press SARL. https://doi.org/10.2991/assehr.k.220504.128. 693-697.

Tongco, M. D. C. (2007). Purposive sampling as a tool for informant selection. *Ethnobotany Research & Applications, 5*, 147-158.

Turing, A. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433-460.

Uslu, F., & Demir, E. (2023). Nitel bir veri toplama tekniği: derinlemesine görüşme. *Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi, 40*(1), 289-299. https://doi.org/10.32600/huefd.1184085

Williams, C. (2007). Research methods. *Journal of Business & Economic Research, 5*(3), 65-72. https://clutejournals.com/index.php/JBER/article/view/2532/2578

Wilson, A., & Anderson, D. (2022). Exploring the confluence of artificial intelligence and creativity: Shaping the future of arts and innovation. *Research Gate.* https://www.researchgate.net/publication/373653388_Exploring_the_Confluence_of_Artificial_Intelligence_and_Creativity_Shaping_the_Future_of_Arts_and_Innovation

Yıldırım, A., & Şimşek, H. (2000). *Sosyal bilimlerde nitel araştırma yöntemleri*. Seçkin Yayıncılık.

Article Information Form

the study does not contain any ethical issues. Consent declaration forms were obtained from all participants, signed and dated, confirming their voluntary participation in the study as seen in the appendices

Plagiarism Statement: This article has been scanned by iThenticate.