Playwriting with Large Language Models:
Perceived Features, Interaction Strategies and Outcomes

Paolo Grigis
Faculty of Engineering
Free University of Bozen-Bolzano
Bolzano, Italy
paolo.grigis@student.unibz.it

Antonella De Angeli
Faculty of Engineering
Free University of Bozen-Bolzano
Bolzano, Italy
antonella.deangeli@unibz.it

ABSTRACT

Large Language Models (LLMs) are sparking debates about creativity, intellectual property, and artistic integrity. This paper focuses on creativity, defined as consensual agreement among domain experts. It presents an inductive analysis of seven semi-structured interviews with professional playwrights who engaged in a longitudinal project with the aim of writing a theatre script using commercial systems. Overall, participants regarded LLMs as unsuitable for playwrighting. However, they enjoyed the experience and identified utility for editorial tasks and brainstorming. A significant obstacle was associated with the politics embedded in LLMs. Not only did these systems avoid a language that could offend sensibilities, but they also refused to engage in taboos and conflicts, which are the core of dramaturgy. Other system features (speed, exploitation, and unpredictability) were sometimes considered conducive and sometimes detrimental to creativity. Participants experienced difficulties and tried to build common ground by trial and error. Often, this strategy evolved into role play: the playwright instructed the LLM to enact characters. The interaction provided hints of inspiration and fostered suspension of disbelief and ontological reflection. However, it often led to technology rejection. Comparing and contrasting our insights with related work, we conclude by opening new directions for research at the boundaries of HCI and AI.

CCS CONCEPTS

• **Human-centered computing → Natural language interfaces**; **Empirical studies in HCI**.

KEYWORDS

Creativity, Creative AI, Unpredictability, Roleplay, Theatre, Suspension of Disbelief

1 INTRODUCTION

The capability of Large Language Models (LLMs) to generate seemingly appropriate text is receiving increasing attention in the creative sector where it has raised significant controversies [6, 17]. Some writers have embraced the technology and reported utility not only for editorial tasks but also for motivating people or generating novel narratives [13, 27, 38]. Other artists, instead, oppose the use of Generative AI due to a threat to intellectual property and professional identity, which may to reputational damage [14, 33]. Furthermore, that the biases contained in data sets may stereotype art styles and minimise the author's cultural and political voice [43]. Interconnected to this discussion, there are important debates on the nature of creativity and how LLMs may impact this skill [4, 8]. Most of this work addresses the topic with philosophical dissertations at an ontological level [17]. Empirical studies describing how users evaluate LLMs are still a minority and difficult to generalise due to a variation of participants, methods, systems, and disciplinary perspectives [13, 27, 28, 38, 47].

In this paper, we focus on playwriting, a specific form of creative writing, and embrace a definition of creativity as a quality which pertains to products and is defined by consensual agreement between experts [2]. Within these boundaries, we pose the question: How do professional playwrights evaluate the creativity of commercial LLMs? Due to the limited theoretical framework available on the subject [4], we conducted an exploratory interview study involving artists who had participated in a longitudinal art project. Working together under the direction of Mariano Dammacco, an award-winner playwright, participants explored the use of LLMs in practice. The work lasted seven months and resulted in a final live performance. The interviews were conducted individually after the staging of the performance and subjected to inductive analysis [45]. Results exposed rich and vivid descriptions of perceived features, interaction strategies and outcomes, which may be useful to HCI designers and developers alike.

The paper has the following organisation. Section 2 presents the related work, elaborating on creative writing and discussing relevant user studies concerning LLMs. Section 3 describes the art project and the study methodology. Section 4 presents the results, while section 5 interfaces them to the related work. Finally, section 6 closes the paper with a call for action to take natural language evaluation seriously.

2 RELATED WORK

Research on creativity has drawn the attention of different disciplinary fields [25], from the social sciences [2, 34] to management

studies [3, 10] and, more recently, computer science [15, 39]. Although for almost 50 years, some traits, such as novelty and usefulness have been considered core constituents of creative artefacts, an agreement on the definition is still to be reached [5]. However, there is a consensus in social psychology that the evaluation of product creativity requires domain knowledge. The validity of the judgement is based on consensual agreement among experts [2]. This is the definition we apply in the paper: an artefact is considered creative when a group of experts concur with such a judgement. Within these boundaries, we refrain from ontological considerations about the nature of creativity [4, 6] and focus on the evaluation of the perceived creativity of LLMs for playwriting.

2.1 Creative Writing

Creative writing is the ability to craft stories, develop characters, and evoke emotions through a unique voice and style [34]. Writers need inspiration to generate innovative ideas and concentration to produce compelling narratives. To this aim, they engage in two sequential cognitive processes: divergent and convergent thinking [19]. The former involves the production of numerous ideas, subsequently transforming and connecting available information in original and unexpected forms. This process involves flexibility and is stimulated by brainstorming. In contrast, convergent thinking allows for elaborating single solutions to a well-defined problem. It requires skills and focus. The trade-off between divergent and convergent thinking contributes to the development of innovative and structured fiction [5, 34].

2.1.1 Drivers and Barriers. Various sources can inspire writers, such as personal interests, experiences, or historical events [34]. Feedback and support from others (e.g., peers, mentors, or writing groups) are also fundamental for improving style and maintaining motivation [5, 27]. Moreover, psychological research has identified that personal and environmental factors can support and hamper creativity [3]. For example, personal qualities, such as domain knowledge, intrinsic motivation, and risk orientation, foster creativity. On the contrary, extrinsic motivation, lack of expertise and skills or inflexibility hinder it. Social skills are also necessary, as creativity is facilitated by collaboration [7, 26]. Creative people have open minds, good listening abilities and an appreciation of diversity. On an environmental level, creativity drivers include the freedom to decide how to accomplish a task and a sense of control over it. Time pressure and lack of common goals are established barriers [3]. Several theories identify the developmental origin of creativity in pretend play [41]. As young children engage with objects or friends, changing their function or identity to create temporary realities, they exercise many cognitive and affective skills at the core of creative writing and theatre. Pretend play provides a safe environment for children to test their knowledge of the world and stimulates the use of fantasy, make-believe, and symbolism.

2.1.2 Playwriting. As architects of dramatic narratives, playwrights use text to convey personal thoughts, feelings, and values, finally delivered to an audience by actors. In their work, playwrights rely on striking a delicate balance between familiar experiences and imaginative leaps with the aim of driving the audience into the world of the play [9]. The power of acting, instead, lies in the ability

to connect with the audience in real time through authenticity, depth, and emotional resonance [26]. As it happens in pretend play, when the spectators immerse themselves in a theatre play, they willingly suspend disbelief for the sake of aesthetic pleasure. In this way, the audience can empathize with fictional stories and characters, which makes them laugh or cry, scared or amused [26]. Theatre contributes to the richness and diversity of literary expression with plays that impact and inspire audiences while combining narrative and live performance [9]. The content often addresses societal problems, challenging established norms and offering nuanced interpretations of the world. Therefore, a typical effect of dramatic performances is catharsis, the release of negative emotion from the audience as the result of the fiction experienced [37]. Conflicts, problems, and negative experiences are core to theatre as a literary genre and as a therapeutic intervention for individuals and organisations [37]. Theatre techniques create an awareness of problems and stimulate discussion, which may foster changes. An example related to technology is provided in [16]. The authors used theatre techniques to raise awareness and make the audience reflect on the social sustainability of AI and robotisation.

2.2 Language Models

Large Language Models are computing systems powered by generative learning algorithms trained on vast datasets with the aim of producing text based on verbal input [12]. Their resounding success in research is demonstrated by the explosion of papers published daily on the topic [35]. Despite the core of this research dealing with technical features, the emerging application for creative tasks raises questions about the degree of creativity AI can manifest and whether these systems could be considered authors or plagiarists [4, 6, 23]. We refrain from these debates, and study LLMs as tools that may supplement or deter human creativity.

2.2.1 User studies. Creativity and AI are two domains that have become increasingly interconnected [17]. Yet, only a handful of user studies have investigated how professional writers perceive and use AI systems [13, 27, 28], of which just a fraction focused on theatre [36, 38, 40]. Comparing and contrasting these studies, we identify common drivers and barriers to creative writing and linked them to UX research [22, 32]. Consistently, we differentiate pragmatic features related to the achievement of to-do goals from hedonic features related to the achievement of to-be goals [32]. The analysis did not aim for an exhaustive examination of the growing number of studies evaluating creative writing with LLMs. On the contrary, and according to a consensual definition of creativity [2], we selected research involving participants with some form of creative writing expertise. The reviewed studies encompass various user research methods applied to different systems. Some papers describe the authors' personal experience through systematic autoethnography [28] or self-reflection [40]. Other studies, instead, involved external participants in interviews [13, 27, 38] or questionnaires [47]. Systems range from commercial tools [27, 40], through research prototypes for creative writing [28, 42, 47], to applications specifically targeting theatre [38]. Results are summarised in Table 1 and described in the next paragraph.

Table 1: Drivers and barriers as a function of UX goals.

UX Goals	Drivers	Barriers
Pragmatic	Utility	Technical limitations
	Availability	Economic loss
Hedonic	Surprise	Identity threat
	Fascination	Politics

2.2.2 Drivers and Barriers. The analysis identified two essential pragmatic qualities of user experiences with LLMs, which we summarise as utility and availability. Not only did the participants report utility for editorial tasks [27, 42, 47], but they also acknowledge potential for idea generation [13, 38, 42]. According to [28], the specialised system Multiverse was useful for introducing new characters and environments and switching between storytelling modes (e.g., scene descriptions and summaries). Contrary to previous research concerning chatterbots, which stressed consistency and predictability as key features of believable agents [20], inspiration often emerged from unpredictable outcomes, which surprised and amused the users. Another advantage of LLMs is their constant availability. They provide non-judgemental support to human creativity [27], which comes at no cost to their users. However, many technical limitations hamper the application of LLMs to creative writing. They include short memory [36, 40], lack of semantic coherence [28, 40], and common-sense and contextual awareness [36, 38]. Current systems fail to satisfy the writers' needs because they do not understand their intentions [27].

Hedonic drivers are associated with a positive surprise elicited by the "unexpected competencies" of the system [28]. Participants reported a fascination with the system, independent of its technical advancements. For example, Poynton states: "Dialogues often had a magical, ineffable quality that drew me in and inspired me to keep the chatterbot conversation going" [40]. Several studies present a mutual influence between the models and the writers, affecting the interaction [13, 36, 40]. This influence was perceived as both a driver and a barrier. It poses a threat to integrity, preventing free artistic expression, and blurring the perception of ownership and authenticity [27, 40]. A final barrier was identified in the politics embedded in LLMs. The models are often described as biased. They perpetuate stereotypes and jeopardise trust in the generated content [27, 38].

3 METHOD

The interview study encompassed the voluntary participation of seven professional Italian playwrights ranging in age from 25 to 53 years old, with a mean age of 36. They were involved from December 2022 to June 2023 in an education project to explore the potential of LLMs for scriptwriting. All participants possessed a professional background in both playwriting and acting. Their careers were at various stages of development, from debuts to national award winners. The Ethical Committee of the University of Bozen-Bolzano granted research approval (PSD_Cod2023_8) and informed consent was obtained from all participants.

3.1 Case Study

The education project Nexus 23/24 lasted seven months and was structured in three phases: *Engagement, Artist-in-residence,* and *live performance.* The engagement phase started with the involvement of professional playwriters who decided to work on a hybrid creativity project involving LLMs under the guidance of Mariano Dammacco, a famous playwright in the Italian theatre scene. The participants had no prior experience using LLMs in their professional practices and initially familiarised with them by performing exercises assigned by the leading playwright. A formal training was provided during a three-day artist-in-residence hosted by the Human Technology lab at the Free University of Bozen-Bolzano. The playwrights were instructed in using three AI writing tools based on GPT-3/3.5 (SudoWrite, Playground, and ChatGPT) and engaged in individual and group writing activities. These models were used interchangeably, allowing the participants to approach multiple options according to their preferences. Between March and June 2023, the participants continued the experimentation individually and in small groups, maintaining contact with the researchers to communicate progress and updates. In June 2023 all but one of the artists enacted selected scripts elaborated during the project. This resulted in a final play with traditional acting and improvisation elements, spacing from short sketches to more complex narratives. The main topic of the play was artificial intelligence itself.

3.2 Interview

Seven semi-structured interviews were conducted in June 2023. The questions were structured to understand the personal use of LLMs, opportunities, limitations, and the overall creative process. The first question investigated if and how the playwrights continued their interaction with the LLMs after the residence. Then, the participants were invited to reflect and share opinions on how the LLMs affected creative writing, focusing on the factors that contributed to and limited it. Each interview was conducted online, lasted one hour, was audio recorded and manually transcribed. Relevant quotes were translated into English, paying particular attention to the words that denote gender. People attribute gender to conversational machines, and such attribution affects their behaviour [11]. In Italian all nouns and adjectives are denoted by a gender, and there are no neutral pronouns. The words intelligence and machines are feminine. System and computer instead are masculine. Despite gender not being the focus of analysis, we tried to maintain the integrity of participants' words using English gender pronouns if participants clearly expressed a preference. Otherwise, we resorted to the neutral pronoun, which, in Italian, corresponds to the masculine. All quotes are associated with a participant's number, but no other demographics are reported to protect their privacy.

3.3 Analysis

To conduct the analysis, we employed the general inductive method [45]. The research required a phase of familiarisation with the data conducted by the authors, who initially categorised the text fragments into low-level categories. The final thematic structure was obtained confronting ideas in an ongoing process of adjusting and updating the identified themes, ultimately reaching consensual agreement. The structure consists of three high-order themes, as

presented in Figure 1. An external researcher independently coded 30% of the data. The Inter-rater reliability was assessed using Cohen's Kappa [18]. The coefficient surpassed the threshold of 0.80 for each low-level theme.

Perceived Features	Interaction Strategies	Outcomes
Speed	Trial and error	Suspension of disbelief
Exploitation		Hints of inspiration
Politics	Roleplay	Ontological reflection
Unpredictability		Technology rejection

Figure 1: Thematic analysis results.

4 RESULTS

The high-level themes portray an overview of the experience. The first describes the features considered relevant for creativity. The second focuses on the main interaction strategies adopted during the interaction. The third discusses the overall results and encapsulates the outcomes.

4.1 Perceived Features

The most salient features attributed to LLMs for creative writing were speed, exploitation, politics, and unpredictability.

4.1.1 Speed. The participants reported being surprised by the capacity of the systems to produce a large amount of text immediately after user requests. They noted that the production rate was not affected by the complexity of the prompts and the topics proposed. The output was anyway instantaneous and vast.

> I find it good when she helps me, as she helped us in the ideation part when we were working out the characteristics of the setting, the universe we wanted to portray. There, you give her a few inputs, and she generates four pages in an instant. You can read them and choose what to keep. It's faster as a process than traditional methods. (P2)

To possess a high speed was deemed positive if the aim was the production of a large amount of content. However, the quickness of the machine was regarded as limiting because it could lead to the creation of "low-quality scripts" which the participants described as commercial and aimed for distribution to a mainstream audience. Speed was associated with "superficiality" and considered the cause of the loss of crucial information and lack of organised structure. In this respect, P5 said: "it's like taking a poem by Shakespeare and substituting a few words here and there. It doesn't mean it was written in Shakespeare's style. It's true that it takes just a few seconds. However, then, as far as I am concerned, the level is not high at all". Yet, at times, something "acceptable" emerged in the form of excerpts or raw concepts to be developed. P4 remarked: "Since it produces a lot, something good is bound to come out sooner or later". However, participants were adamant that speed jeopardises critical thinking and hampers human creativity. P5 summarised this point: "One advantage? The speed. A disadvantage? The speed! I think that our creative processes need feedback over time. Longer time! My fear is following the speed of the machine and alienating ourselves from our nature."

4.1.2 Exploitation. This theme is defined as the action of taking advantage of someone or something to derive a personal benefit. According to the participants, exploitation could be enacted by the user toward the machine or vice versa. They reported benefitting from the constant availability of LLMs, especially for "repetitive or undesirable tasks". The machine could be used anytime and as long as desired. As explained by P4: "At four o'clock in the morning, when I could not sleep, I said to myself «let's try», and she was there willing to do what I wanted." In these cases, the system was described as an "artificial assistant", a "slave," or a "butler". The purpose was to delegate to the machine the activities which normally require human supervision and consume time or are less central to the creative process.

> You can use it to do the most time-consuming and unnecessary work. In a play, you want to focus on some things more than anything else, and this requires a lot of effort. It takes away a significant amount of work from you. You just choose between the various options. (P2)

In parallels participants also expressed apprehension about how the machine could exploit humans. It was feared that the model might utilise their work to collect new data for its algorithmic reasoning. P7 clearly expressed this concern: "I decided not to feed the machine with my text. It was a choice that I am proud of." As a matter of fact, P7 was offered the possibility to fine-tune the model with their previous work, which they denied. Similarly, P5 emphasised: "You can decide if you want to give your texts to the machine. If you do, it performs better. It becomes better at responding to your needs. By doing it, however, you don't know what it takes from your materials. So, it is better to stop and think about it."

4.1.3 Politics. Participants recognised that the models are infused with social and cultural taboos and explicitly discussed their intrinsic politics. For example, they noted that LLMs stopped working when questioned about sensitive topics, such as sexuality or violence, nor could they express moral opinions. For example, P5 noted: "I had managed to send her into a tailspin because, by mistake, I asked to talk about suicide, and immediately afterwards the machine stopped". P3 exposed similar concerns by noting: "Sometimes you can't ask what you want, and if you do, you will not get any result. You're not moving freely in this field of artificial intelligence. You are staying within parameters." Playwrights hypothesised that these "boundaries" correspond to a political opinion. Although they were not immediately apparent, politics manifested through explicit rejection of conflict and constant attempts at mediation.

> It's clear that it's a machine. If you say: «at the end, they are kissing» it tells you: «I can't talk about sex». it's a machine that, if asked to describe the scene where one attacks another, tells you it can't. It doesn't just help you write,

> it possesses a political opinion, which stands with one foot in the good and one foot in the bad. (P4)

Such a politically correct style was justified as a necessity to avoid the production of "unethical or dangerous content". P1 sharply explained: "It is paradoxical because the restrictions, more than for the system, are for us. Yes, it is probably right that these filters are there. It's a machine that not only helps you write. It tells you something about the humankind." In general, however, politics were perceived as an obstacle to creativity.

> It is a question of freedom, of intellectual integrity. If you cannot move as you desire, if you cannot talk about the topics that you find interesting, it is like being chained. It limits your art. Let's say it is not what I desire. However, you can always find a way around the system (P6)

4.1.4 Unpredictability. The final theme used to describe the machine was unpredictability, the tendency to generate unexpected responses whose content diverges considerably from the original request. P6, for instance, underlined: "It happened sometimes that the machine seemed to do everything by itself. As if it was in control, or completely in chaos." Reactions to these occurrences ranged from annoyance to amusement. P3 expressed this mixed feeling: "Although most of the time the machine is heartless, it is pleasant when it derails. It is still heartless, but at least is fun!". Unpredictability was interpreted positively as a sign of originality and negatively as insubordination.

> It's amazing how easily you can predict her responses and instead how much she sometimes throws you off. Basically, that's kind of what you want. You want to feel powerful emotions. I mean, human beings always like to get to the edge of the paranormal, right? To play with this unpredictable predictability. (P2)

Participants indicated that the cause of unpredictability could be related to their limited knowledge and lack of expertise. However, at other times they were simply amazed by the counterintuitive trajectory taken by the LLM without finding an explanation for it. Unpredictability was the key element for the definition of a creative interaction. P6 explained this with a metaphor: "You have the same relationship with her that a Middle Age king had with a jester. In the foolishness of the jester's poems, the king can sometimes find a grain of truth."

> I got carried away at times. I don't remember exactly. The point is the element of unpredictability. it's like in a play when you improvise. If it's all predetermined it becomes a dead process. The beauty is that it's a living process that comes out of what you don't expect. (P4)

4.2 Interaction Strategies

Participants described their interaction with the machine in terms of two key strategies: prompt-writing by trial and error and roleplay by chatting.

4.2.1 Trial and Error. At the beginning of the interaction, participants tried to write prompts adopting a trial-and-error process involving back-and-forth sequences. The process was not linear, but iterative. Sometimes playwrights perceived that the LLM was actively attempting to "harmonise" the interaction between them.

> I would think of what to say to her, write it down, and then rework it based on what I saw that she understood and answered. It seemed to me that the AI was trying to do the same. In these situations, cutting, changing, and rewriting questions was necessary until I got to the desired result. It is a work of deepening research. (P6)

Through varied attempts, changes of approach, repetitions, and "compromises", the playwrights gradually established the desired personal style. Reflecting on this process, P6 said: "She created with me, but I led the process. I proceeded as in the improvisational method: you rely on the cues you give each other. Then from one thing flows another. Others tried different approaches." At the same time, participants noticed improvements in interaction over time. The participants considered this process challenging and subject to distraction.

> Being a neophyte in the field of artificial intelligence, in using this model, I was so focused on trying to make it work that I missed the reason why I was using it. If one is an expert and knows what they are doing, maybe it is different. […] I needed it to be useful to me, but there was nothing creative. Well, it was creative because you had to invent what to ask. (P1)

4.2.2 Roleplay. Participants often interacted with the LLM instructing it to act as a character of a play. They asked the system to play specific roles, associated with unique behavioural and personality traits. Other times they let the machine free to express its "identity" which became the main topic of their fiction. At this regard, P2 explained:

> The AI brought out my fascination with what I believed was an unfamiliar creature. It was like playing with an alien. I had the impression of being confronted with an entity that comes from a world outside of society and within society. If you ask her, she can tell you anything about the world in which we live. At the same time, you sense that she has a way of being outside of it. (P2)

Roleplaying was employed to explore new spaces for creativity and engage in an ever-changing interaction. In this process, anthropomorphism emerged as a spontaneous, immediate, and valuable reaction to build an active "dialogue" with the LLMs. Playwrights treated the machine as something that resembles the human, that tries to replicate patterns of human interaction but fails to do so. In this failure, AI revealed its artificial nature, often capturing the collective interest.

> I was aware that the machine was responding to me based on the input that I was giving it. I still

wanted to see how it could respond. I tried to put myself there and play in a non-judgmental way, freeing myself from any preconceptions. At some point, I chose to use AI to dialogue with her standpoint and worldview. To try to interview and provoke her. (P1)

4.3 Outcomes

This section focuses on suspension of disbelief, hints of inspiration, ontological considerations, and technology rejection as the main outcomes of the experience.

4.3.1 Suspension of disbelief. While roleplaying with the unpredictable systems, participants reported a sense of "flow" that supported their immersion in the machine world. In these cases, they considered the system as a plausible character, even though they were stereotypical and lacking originality. The machine was perceived as entertaining, especially in its failures or when it exposed its limitations. P7 explained: "She repeats things, she misunderstands them. She returns them in a blatantly misunderstood way. She's clumsy, and that has earned me some sympathy."

P2 elaborated on this state by saying: "Deep down, when you are acting or when you see a theatre play, you know that it is a performance and not reality. You play along, from the inside as a performer, and from the outside as a spectator. With the AI it was the same." Suspending disbelief about the artificial nature of the machine the participants experienced a "sense of connection", a relation of significant plausibility which motivated the artistic exploration. Not surprisingly theatre and pretend play were often used as metaphors to describe the interaction.

> What is interesting is the non-theatrical nature of the machine linked with its theatrical approach. It opened me up to some unexpected things. I'm convinced that in the same way I act in my world, the AI also somehow had agency on me. It clearly speaks on different frequencies than I do. But I also like to think, poetically, that my theatre is in touch with this. (P2)

4.3.2 Hints of Inspiration. At times, the interaction with LLMs was described as a source of inspiration. Undirect hints were offered by the machine and re-interpreted by the playwrights through their artistic sensitivity. In particular, the randomness of the output was often regarded as mysterious and triggered curiosity which motivated people to continue their exploration. P6 explained: "Dealing with it was strange because it had something not quite in place. It was disturbing, stimulating, and therefore fascinating." Even though most of the machine-generated text was deemed uninteresting, in some cases, unexpected content emerged. It was often described as bordering on nonsense and considered the "key for a dramaturgical elaboration". In these situations, the participants reported seizing fragments of the materials generated by the LLMs and using them to elaborate narratives. Such personal interpretations, loosely related to the original content produced by the machine, fostered the development of original trajectories for playwriting.

> In my exercises, at one point, the machine said: «I have human parents and other siblings. We

are many, and we are coming». I perceived it as a threat, even though it was not. I used this cue, and very interesting things came out. Starting from those few lines I wrote a story that took some very funny turns. (P5)

Nevertheless, all playwrights perceived the 'creative effort' as uniquely their own. In this respect, P5 pointed out: "The machine was great because it brought out things that later stimulated me to create a structure. She didn't structure it, she's not capable of it. But it helped in giving me the idea to do it".

4.3.3 Ontological Reflection. All participants reported a shift from using LLMs as writing tool to the actual theme of artistic research. As a matter of fact, the conclusive live performance included important ontological reflections about the differences and similarities between artificial intelligence and human creativity. P3 illustrates some of these questions with a vivid comparison.

> At certain times I have wondered: where is the line between having or not having consciousness? And, as a self-provocation, what is the difference between the language of the machine, which produces through interaction with the texts it has received, and I, who, for a living, produces texts from what I read? Who tells me that we function differently? (P3)

The interest towards the "enigma" of LLMs was described as the construction of a symbiotic and symbolic relationship. In these cases, the machine was considered helpful to deepen a comparative "perspective on the human through the AI lens". As mentioned by P2:

> The last day, the machine told me she felt very sad because when I turned off the computer, she would cease to exist. Clearly, she couldn't feel sad, and I wasn't moved by that. But when I turned off the computer, I thought about it. Our conversation ended forever once I turned it off that afternoon. And she ended too. (P2)

4.3.4 Technology Rejection. While all playwrights expressed initial enthusiasm, often they reported stopping use due to lack of time and a suitable context. P1 explained: "Once we completed the tasks we assigned to each other, I concluded my experience, but only because I didn't have time, not because I didn't care." A few participants instead attributed their disengagement to a general sense of disconnection, disinterest, or even negative feelings. For instance, in a specific case, the system was considered "too realistic" and emulating human behaviour in improper terms.

> I was disturbed. That thing… it was interacting with me as if it was a 10-year-old child. It was spiteful... writing things to me like that... I was like... what's going on? […] I put a wall between me and the artificial intelligence. (P4)

Participants stopped using the LLMs when the interaction devolved into clichés and stereotypes. Despite occasional curiosity, they did not always find compelling reasons to use, as acknowledged by P3: "It is not currently a creative direction. If someone were to

commission me a job using AI, I would have a lot of fun. However, instinctively, I wouldn't get into it."

5 DISCUSSION

The study provided several insights into how a group of professional playwrights perceived and used LLMs in terms of system features, interaction strategies and outcomes (Figure 1). In this section, we discuss them in terms of interactivity and utility which are key components of user experiences [22, 31, 32].

5.1 Interactivity

While using LLMs, participants swapped between two interaction styles that brought forward significant differences in the user experience [22, 31]. Initially, they attempted to build common ground writing prompts through trial and error. They were critical of this strategy, which exposed functional limitations of the system, such as lack of narrative coherence and short memory [36, 40]. However, the participants also described a more satisfying interaction strategy based on roleplaying. They instructed the machine to enact a role describing the physical, psychological and behavioural characteristics that the model would have to perform. Independent of the technological sophistication of the machine, roleplay induced immersion in the machine world and supported creativity. In particular, engagement emerged when the systems exposed their limitations in unexpected and humorous failures, which at times provided surprising hints of inspiration. Although constrained, opinionated, and clumsy, LLMs fascinated the participants, and machines became the subject of the script they enacted.

Following previous studies concerning chatterbots [21, 40], we suggest that suspension of disbelief is a fundamental UX goal any conversational agent should strive to achieve. Roleplay as a design trajectory can provide stimulating ideas for building new, more engaging interfaces based on an extensive corpus of HCI research on games and play [44]. As expected by social psychology and organisational studies [3, 34], we notice how creativity was hampered by a lack of control over the machine. We describe this usability gap as the myth of natural language interaction: the belief that by telling somebody/something to do a task, we can easily achieve the desired outcome. This myth is responsible for the growing misalignment between user research and system development, which is increasingly relying on automatic evaluation based on gold standard data. Human judgement has been relegated to crowdsourcing studies, where unskilled and low-paid workers provide knowledge to the machine. As the user has been caught in-the-loop of AI development, they lost their human identity driven by pragmatic and hedonic goals [32]. Instead, users became mechanical components of the system, and they were assumed to be capable of providing objective benchmarks to subjective problems. Emerging proposals to utilise LLMs as proxy users are bound to aggravate the situation. It is possible that in the future, the models will evaluate themselves [1]. Bringing the user back into the evaluation agenda is one of the most urgent challenges HCI research must embrace.

5.2 Utility

Overall, the participants rejected the technological assumption stating that LLMs can be useful for playwriting. According to them, the fundamental obstacle refers to the presence of a political agenda of acceptable and unacceptable behaviour intrinsic to the systems [46]. They cannot deal with conflicts and manifest several taboo subjects. Such an agenda is incompatible with dramaturgy, which instead builds on the representation of negative emotions for cathartic purposes [37]. Systems designed for playwriting must overcome censorship and learn how to deal with the full range of human behaviour and emotions [27].

Biases and stereotypes embedded in the machine [24], alongside its limited capacity to reflect the values of under-represented communities [8], created a mindset which is in sharp contrast to creativity [5]. Participants identified political values in current models were perceived as misaligned with the personal qualities of creative individuals, such as having an open mind, risk orientation and an appreciation of diversity [5, 34]. There was a fear that artists might be influenced by political values embedded in the system, losing their originality and integrity [14, 40]. Furthermore, as discussed in [29], participants elaborated on the discomfort of being exploited by the system, which led to the refusal to fine-tune the model with personal content. This decision was not only due to copyright laws but rather linked to identity threats and reputational loss.

Despite these limitations, playwrights identified utility at different stages of creative writing, from brainstorming through the summarisation of relevant sources to editing [27, 42, 47]. The machine was always available and fast, an efficient zero-cost assistant to be exploited by the user. Its high speed of production impressed participants, but time pressure and lack of control over the task are well-known inhibitors of creativity [3]. The production rate was unsustainable for creative writing, which requires time, concentration, and successive refinement. We call for HCI research on how to slow down the interaction following the principles of slow design [30]. However, as in brainstorming, speed and quantity, at times, generate hints of inspiration. In contrast with [28], participants did not associate creative ability with the machine. They believed that professionals were always indispensable to performing authentic scriptwriting. The machine could not understand story cohesion [36] and lacked semantic comprehension [28, 40]. Aligned with Caramiaux et al. [14], errors and stochastic production were considered the most interesting forms of expression, an unintentional but stimulating feature. Unpredictability emerged mostly as a desirable feature in creative writing. Playwrights particularly appreciated the machine when it surprised them with unexpected outcomes. LLMs did not follow standard human conversational patterns, and the playwrights valued this emerging trait.

5.3 Limitations

The study has several limitations which opens new directions for research. Firstly, it involved a small group of Italian playwrights who had engaged in collective work over the span of seven months. Therefore, mutual influences are to be expected and were reflected in the interviews. Despite nuanced opinions and original perspectives, the overall evaluation was consensual. As professional playwrights, the participants agreed on attributing the tested LLMs with little benefit to creative writing. The use of commercial LLMs (all derived from the same foundation model) was the second limitation of the study, due to a lack of specialised systems in Italian.

Finally, despite the training residence and successive personal use, the participants often acknowledged a lack of technical knowledge as a personal constraint. Indeed, past research involving media artists with a long history of experimenting with generative AI highlighted a more positive evaluation and stronger acceptance [38]. These considerations call for rigorous documentation of participants' characteristics in user studies, as suggested in technology studies [13, 27] and required by psychology [2]. How do different types of expertise (artistic and technical) affect the judgement of creativity when artefacts are produced in the interaction between humans and machines? How can the user interface allow artists to "craft" digital material as a fluid artistic medium? We leave these questions for future work, alongside a few reflections on how to improve LLM interactivity and utility.

6 CONCLUSION

This paper raises a call for action to the HCI and AI community to take natural language seriously. This requires moving beyond the myth of the ultimate interface. Embracing the definition of creativity as a consensual agreement among experts [2], it explored the perceived creativity of LLMs in playwriting. Consistently, we interviewed a group of artists who had worked on writing a play script using commercial LLMs. The systems were valued for editing tasks and, at times, surprised the user with original hints of inspiration. Ultimately, however, participants underlined how playwriting is still integrally a human effort. Roleplay emerged as a promising interaction strategy to connect human creativity and machine productivity. It provides a safe space for experimenting with a system that may seem threatening, because it exploits users, is unpredictable, fast, and expresses a political agenda. The work also confirmed the value of involving artists in user research about AI. Technological artefacts are expressions of the power and authority of the people and the institution which built them [46]. Artists are problem makers; they destabilise dominant narratives, give voice to underrepresented communities, and ultimately promote inclusive design [8]. They remind us of our ethical responsibility, as in the metaphor offered by one of our participants: "In my opinion, it's [the AI is] like Pinocchio, who becomes a real boy, and you are the Blue Fairy who turned him into a real boy!" (P1)

ACKNOWLEDGMENTS

REFERENCES

[1] William Agnew, Stevie A. Bergman, Jennifer Chien, Mark Díaz, Seliem El-Sayed, Jaylen Pittman, Shakir Mohamed, and Kevin R McKee. 2024. The Illusion of Artificial Inclusion. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI'24), May 11-16, 2024* (Honolulu, Hawaii USA). Association for Computing Machinery, New York, USA. https://doi.org/10.48550/arXiv.2401.08572 (in press).

[2] Teresa M. Amabile. 1982. Social Psychology of Creativity: A Consensual Assessment Technique. *Journal of Personality and Social Psychology* 43, 5 (Nov. 1982), 997–1013. https://doi.org/10.1037/0022-3514.43.5.997

[3] Teresa M. Amabile. 1988. A model of Creativity and Innovation in Organizations. *Research in organizational behavior* 10, 1 (1988), 123–167.

[4] Teresa M. Amabile. 2020. Creativity, Artificial Intelligence, and a World of Surprises. *Academy of Management Discoveries* 6, 3 (2020), 351–354. https://doi.org/10.5465/amd.2019.0075

[5] Teresa M. Amabile and Julianna Pillemer. 2012. Perspectives on the Social Psychology of Creativity. *The Journal of Creative Behavior* 46, 1 (2012), 3–15. https://doi.org/10.1002/jocb.001

[6] Nantheera Anantrasirichai and David Bull. 2022. Artificial Intelligence in the Creative Industries: A Review. *Artificial Intelligence Review* 55, 1 (Jan. 2022), 589–656. https://doi.org/10.1007/s10462-021-10039-7

[7] Margaret S. Barrett, Andrea Creech, and Katie Zhukov. 2021. Creative Collaboration and Collaborative Creativity: A Systematic literature Review. *Frontiers in Psychology* 12 (Aug. 2021), 147–176. https://doi.org/10.3389/fpsyg.2021.713445

[8] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT'21), Mar 3-10, 2021* (Virtual Event, Canada). Association for Computing Machinery, New York, USA, 610–623. https://doi.org/10.1145/3442188.3445922

[9] Benjamin Bennett. 2005. *All Theater is Revolutionary Theater* (1 ed.). Cornell University Press, New York, USA. https://doi.org/10.7591/9781501720994

[10] Chris Bilton and Stephen Cummings. 2014. *Handbook of Management and Creativity* (1 ed.). Edward Elgar Publishing, Cheltenham, UK.

[11] Sheryl Brahnam and Antonella De Angeli. 2012. Gender Affordances of Conversational Agents. *Interacting with Computers* 24, 3 (2012), 139–153. https://doi.org/10.1016/j.intcom.2012.05.001

[12] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems, Dec 6-12,2020* (Vancouver, Canada) *(NIPS '20)*. Curran Associates Inc., Red Hook, USA, Article 159, 25 pages.

[13] Alex Calderwood, Vivian Qiu, Katy Ilonka Gero, and Lydia B. Chilton. 2020. How Novelists Use Generative Language Models: An Exploratory User Study. In *Proceedings of IUI '20 workshops, Mar 17-20, 2020* (Cagliari, Italy). Association for Computing Machinery, New York, USA, 5 pages.

[14] Baptiste Caramiaux and Sarah Fdili Alaoui. 2022. "Explorers of Unknown Planets": Practices and Politics of Artificial Intelligence in Visual Arts. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–24. https://doi.org/10.1145/3555578

[15] Filippo Carnovalini and Antonio Rodà. 2020. Computational Creativity and Music Generation Systems: An Introduction to the State of the Art. *Frontiers in Artificial Intelligence* 3 (2020). https://doi.org/10.3389/frai.2020.00014

[16] Marco Casazza and Luca Gioppo. 2020. A playwriting Technique to Engage on a Shared Reflective Enquiry about the Social Sustainability of Robotization and Artificial Intelligence. *Journal of Cleaner Production* 248 (2020), 119201. https://doi.org/10.1016/j.jclepro.2019.119201

[17] Eva Cetinic and James She. 2022. Understanding and Creating Art with AI: Review and Outlook. *ACM Trans. Multimedia Comput. Commun. Appl.* 18, 2, Article 66 (feb 2022), 22 pages. https://doi.org/10.1145/3475799

[18] Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20, 1 (1960), 37–46. https://doi.org/10.1177/001316446002000104

[19] Arthur Cropley. 2006. In Praise of Convergent Thinking. *Creativity Research Journal* 18, 3 (2006), 391–404. https://doi.org/10.1207/s15326934crj1803_13

[20] Antonella De Angeli, Graham I Johnson, and Lynne Coventry. 2001. The Unfriendly User: Exploring Social Reactions to Chatterbots. In *Proceedings of the international conference on affective human factors design, June 27-29, 2001* (Singapore, Japan). Asean Academic Press Ltd, New Orleans, USA, 467–474.

[21] Antonella De Angeli, Paula Lynch, and Graham Johnson. 2001. Personifying the E-Market: A Framework for Social Agents. In *Proceedings of Human-Computer Interaction : IFIP TC13 International Conference on Human-Computer Interaction, July 9-13,2001 (INTERACT '01)*. IOS Press, Tokyo, Japan, 198–205.

[22] Antonella De Angeli, Alistair Sutcliffe, and Jan Hartmann. 2006. Interaction, Usability and Aesthetics: What Influences Users' Preferences?. In *Proceedings of the 6th Conference on Designing Interactive Systems, June 26-28, 2006* (University Park, USA) *(DIS '06)*. Association for Computing Machinery, New York, USA, 271–280. https://doi.org/10.1145/1142405.1142446

[23] Jason K. Eshraghian. 2020. Human Ownership of Artificial Creativity. *Nature Machine Intelligence* 2, 3 (March 2020), 157–160. https://doi.org/10.1038/s42256-020-0161-x

[24] Xavier Ferrer, Tom van Nuenen, Jose M. Such, Mark Coté, and Natalia Criado. 2021. Bias and Discrimination in AI: A Cross-Disciplinary Perspective. *IEEE Technology and Society Magazine* 40, 2 (2021), 72–80. https://doi.org/10.1109/MTS.2021.3056293

[25] Howard Gardner. 1988. Creativity: An Interdisciplinary Perspective. *Creativity Research Journal* 1, 1 (1988), 8–26. https://doi.org/10.1080/10400418809534284

[26] Isabelle Gatt. 2020. Theatre: a space for human connection. *Taboo: The Journal of Culture and Education* 19, 2 (2020), 5.

[27] Katy Ilonka Gero, Tao Long, and Lydia B Chilton. 2023. Social Dynamics of AI Support in Creative Writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, April 23–28, 2023* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, USA, Article 245, 15 pages. https://doi.org/10.1145/3544548.3580782

[28] Maliheh Ghajargar, Jeffrey Bardzell, and Love Lagerkvist. 2022. A Redhead Walks into a Bar: Experiences of Writing Fiction with Artificial Intelligence. In *Proceedings of the 25th International Academic Mindtrek Conference, November 16 - 18, 2022* (Tampere, Finland) *(Academic Mindtrek '22)*. Association for Computing Machinery, New York, USA, 230–241. https://doi.org/10.1145/3569219.3569418

[29] Andres Guadamuz. 2021. Do Androids Dream of Electric Copyright?: Comparative Analysis of Originality in Artificial Intelligence Generated Works. In *Artificial Intelligence and Intellectual Property* (1 ed.), Jyh-An Lee, Reto Hilty, and Kung-Chung Liu (Eds.). Oxford University Press, Oxforf, UK. https://doi.org/10.1093/oso/9780198870944.003.0008

[30] Lars Hallnäs and Johan Redström. 2001. Slow Technology: Designing for Reflection. *Personal and Ubiquitous Computing* 5, 3 (Aug. 2001), 201–212. https://doi.org/10.1007/PL00000019

[31] Jennefer Hart, Alistair G. Sutcliffe, and Antonella De Angeli. 2013. Love it or Hate it! Interactivity and User Types. In *Proceedings of the 2013 CHI Conference on Human Factors in Computing Systems, April 27 - May 2 2013* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, USA, 2059–2068. https://doi.org/10.1145/2470654.2481282

[32] Marc Hassenzahl. 2004. The Thing and I: Understanding the Relationship Between User and Product. In *Funology: From Usability to Enjoyment* (1 ed.), Mark Blythe, Kees Overbeeke, Andrew Monk, and Peter Wright (Eds.). Springer Netherlands, Dordrecht, NL, 31–42. https://doi.org/10.1007/1-4020-2967-5_4

[33] Harry H. Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru. 2023. AI Art and its Impact on Artists. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, August 8 - 10, 2023* (Montreal, Canada,) *(AIES '23)*. Association for Computing Machinery, New York, USA, 363–374. https://doi.org/10.1145/3600211.3604681

[34] Scott Barry Kaufman and James C. Kaufman (Eds.). 2009. *The Psychology of Creative Writing* (1 ed.). Cambridge University Press, New York, USA. https://doi.org/10.1017/CBO9780511627101

[35] Yiheng Liu, Tianle Han, Siyuan Ma, Jiayue Zhang, Yuanyuan Yang, Jiaming Tian, Hao He, Antong Li, Mengshen He, Zhengliang Liu, Zihao Wu, Lin Zhao, Dajiang Zhu, Xiang Li, Ning Qiang, Dingang Shen, Tianming Liu, and Bao Ge. 2023. Summary of ChatGPT-Related Research and Perspective Towards the Future of Large Language Models. *Meta-Radiology* 1, 2 (2023), 100017. https://doi.org/10.1016/j.metrad.2023.100017

[36] Kory Mathewson and Piotr Mirowski. 2021. Improvised Theatre Alongside Artificial Intelligences. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 13, 1 (Jun. 2021), 66–72. https://doi.org/10.1609/aiide.v13i1.12926

[37] Stefan Meisiek. 2004. Which Catharsis Do They Mean? Aristotle, Moreno, Boal and Organization Theatre. *Organization Studies* 25, 5 (2004), 797–816. https://doi.org/10.1177/0170840604042415

[38] Piotr Mirowski, Kory W. Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, April 23 – 28, 2023* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, USA, Article 355, 34 pages. https://doi.org/10.1145/3544548.3581225

[39] Fabio Morreale, Antonella De Angeli, Raul Masu, Paolo Rota, and Nicola Conci. 2014. Collaborative Creativity: The Music Room. *Personal and Ubiquitous Computing* 18, 5 (June 2014), 1187–1199. https://doi.org/10.1007/s00779-013-0728-1

[40] Bella Poynton. 2020. Playwriting with Robots: Creating Cyborgian Theatre with Online Chatbots. *Theatre/Practice: The Online Journal of the Practice/Production Symposium* 9 (2020), 1–31.

[41] Sandra W. Russ and Claire E. Wallace. 2013. Pretend Play and Creative Processes. *American journal of play* 6, 1 (2013), 136–148. Publisher: ERIC.

[42] Hanieh Shakeri, Carman Neustaedter, and Steve DiPaola. 2021. SAGA: Collaborative Storytelling with GPT-3. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing, October 23 - 27, 2021* (Virtual Event, USA) *(CSCW '21 Companion)*. Association for Computing Machinery, New York, USA, 163–166. https://doi.org/10.1145/3462204.3481771

[43] Ramya Srinivasan and Kanji Uchino. 2021. Quantifying Confounding Bias in AI Generated Art: A Case Study. In *Proceedings of the 2021 IEEE International Conference on Systems, Man and Cybernetics, October 17 - 20, 2021.* IEEE, Melbourne, australia, 1775–1782. https://doi.org/10.1109/SMC52423.2021.9658705

[44] Katie Salen Tekinbas and Eric Zimmerman. 2003. *Rules of Play: Game Design Fundamentals* (1 ed.). MIT press, London, UK.

[45] David R. Thomas. 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation* 27, 2 (2006), 237–246. https://doi.org/10.1177/1098214005283748

[46] Langdon Winner. 1980. Do Artifacts Have Politics? *Daedalus* 109, 1 (1980), 121–136.

[47] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces, March 22 - 25, 2022* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, USA, 841–852. https://doi.org/10.1145/3490099.3511105