The Impact of AI on Human Creativity: An Experimental Approach

Bachelor Thesis

Author: Steven Mkhitarian (Student ID: 7379848)
Supervisor: Univ.-Prof. Dr. Markus Weinmann
Co-Supervisor: Sercan Demir

Department of Business Analytics
Faculty of Management, Economics and Social Sciences
University of Cologne

April 17, 2024

Abstract

This study delves deeply into the impact of feedback from Language Learning Models (LLMs) on the quality and originality of human creative processes, particularly as Artificial Intelligence (AI) becomes increasingly integrated into creative pursuits. As AI assumes a larger role in creative tasks, it becomes crucial to evaluate how AI-driven feedback influences creativity. This study aims to build upon existing literature by exploring the nuanced effects of various feedback techniques on creative outputs. An experiment was conducted using creative tasks to compare and analyze the effects of specific, generic, and no feedback modalities delivered by the LLM ChatGPT. Employing a detailed experimental approach alongside a mix of quantitative and qualitative techniques, the study seeks to provide empirical insights into how AI feedback impacts creativity. The findings reveal considerable differences across feedback modalities and types of creative tasks, highlighting significant implications for the application of AI in creative contexts. The use of AI in enhancing creative activities notably demonstrates its potential to foster creativity growth, suggesting expansive possibilities for AI's role in creative industries.

Contents

List of Tables

1 Introduction

Artificial Intelligence (AI), with a specific focus on large language models (LLMs) such as ChatGPT, has significantly enhanced human efficiency in various areas of our lives (Abdullah, Madain, & Jararweh, 2022). This advancement is most evident when there is a collaboration between humans and AI, particularly through prompt engineering, positioning AI not merely as a ghostwriter but as a robust tool for innovation (Chen & Chan, 2023). The widespread integration of AI into various parts of human life signals a new age of innovation and transformation, simplifying activities from the mundane to the complex and facilitating decision-making processes. (Dwivedi et al., 2019). Many studies show that AI's influence extends far beyond boosting workplace productivity.AI profoundly impacts a wide array of human activities, from performing everyday tasks to facilitating complex decision-making in healthcare, economy, and beyond (Jiang et al., 2017; Furman & Seamans, 2018). The use of AI has led researchers to predict a paradigm shift that could significantly impact humanity, especially considering the rapid pace of AI development and improvement (Dwivedi et al., 2019). In light of these developments, the exploration of AI's impact on human creativity emerges as a compelling and timely endeavor. As AI systems become increasingly sophisticated and popular, understanding their influence on human creative activities becomes imperative for both theoretical understanding and practical applications (Anantrasirichai & Bull, 2020). Continued research is particularly crucial given the rapid advancement and improvement of LLMs; as on a consistent basis, new breakthrough or groundbreaking technology is being developed (Dwivedi et al., 2019), turning the focus next generation of ChatGPT. Today, generative AIs have the capacity to create a wide array of outputs, including highly photorealistic photos (like Midjourney), videos (like Sora), and more (Marchesi, 2017; Liu, Huang, Yu, Wang, & Mallya, 2020); opening up vast new scenarios of possibilities for innovation and advancement in various fields, while at the same time underlining the urgency and significance of understanding the potential of AI in creative pursuits (Anantrasirichai & Bull, 2020).

In the course of this technological transformation, the link between AI and human creativity emerges as an attractive topic of research. Creativity, a vital component of human intelligence, has long been seen as a distinguishing human characteristic (Norton, Heath, & Ventura, 2013). However, the rising capabilities of AI challenge this traditional consensus, especially when AI systems display incredible creative potentials, from producing art to offering nuanced feedback in creative processes (Cetinic & She, 2021; Chen & Chan, 2023). This exploration is especially relevant due to the swift advancements in AI, including breakthroughs in generative technologies that can create highly realistic images and texts (Marchesi, 2017). These developments have ignited debates about how much AI can enhance, or perhaps even exceed, human creativity. Notably, cases where AI has won in creative competitions with professional artists highlight the urgent need to understand AI's role in creative fields (Anantrasirichai & Bull, 2020). The impact of AI on human creativity has the potential to provide meaningful findings at the intersection of technology, cognition, and innovation. As AI systems become more complex and pervasive, understanding their influence on creative activities is critical for both theoretical and practical applications (Atkinson & Barker, 2023) The successful use of AI often depends on its creative settings, for instance, the study Chen and Chan has demonstrated that using Language Model Machines (LLMs) as sounding boards can significantly enhance the creative performance of non-experts whereas utilizing LLMs as ghostwriters can even result into worse results (Chen & Chan, 2023).

Despite the growing interest in the role of AI within creative processes, there is still further need for empirical research concerning the impact of AI-generated feedback on human creativity, particularly considering variation of feedback. Early studies have indicated the potential of AI, especially LLMs, to enhance creative outputs e.g creative writing where the LLM helps mitigate the "writers block" and encourages a faster start in the process (Yuan, Coenen, Reif, & Ippolito, 2022). However, the relationship between the type of feedback and specific creative outcomes remains undefined. Chen and Chan (Chen & Chan, 2023) demonstrated that a sounding board modality is more effective than a ghostwriter modality, however the optimal type of input for the sounding board modality to achieve ideal outcomes is still unknown, leading to believe that the nature of feedback and the potential differences among various types of input, each characterized by distinct data levels, are not well understood.

The aim of this research is to bridge this gap by analysing the extent to which different feedback modalities from AI affect creative outcomes and semantic quality; by answering to what extent the variation of feedback provided by LLMs influences the quality and creativity of human work. Answering this question will contribute to a deeper understanding of the complex interplay between AI and human creativity in the evolving landscape of innovation; thus acknowledging the perspectives of Barile et al. on the transition from mechanical visions of AI to systemic visions of Intelligence Augmentation (IA) and its impact on human creative processes (Barile, Bassano, Piciocchi, Vito, & Spohrer, 2022). The primary hypothesis discussed in this study is that detailed, contextually rich feedback from AI, as compared to more generic or no feedback, significantly improves the quality and creativity of human creative outputs.

To validate this idea, an experimental technique was utilized. Participants participating in creative writing activities responded to varied feedback conditions given by ChatGPT. This research uses a mixed-methods approach to systematically investigate the link between AI-generated feedback and creative production, including quantitative and qualitative assessments. It also intends to guide future AI integrations in creative domains by undertaking a thorough evaluation of AI feedback's influence on semantic creativity and quality, therefore expanding the symbiotic potential of human-machine collaboration and promoting innovation in creative activities.

2 Literature Review

Human creativity has fueled innovation and progress throughout history, shaping societies and the world itself. From the earliest cave paintings to the technological advancements of the modern age, the human capacity for creativity has led to significant cultural, scientific, and technological breakthroughs (Pope, 2005). Our understanding of creativity has evolved significantly over time, reflecting changing values, philosophies, and scientific knowledge and discoveries (Albert & Runco, 1998). Historically, creativity was often seen as divinely inspired and a trait reserved only to humans (Dacey, 2011). This outlook, however, evolved over time, giving rise to a more sophisticated comprehension that acknowledges the interplay of psychological, biological, and sociocultural influences (Amabile & Pillemer, 2012). Modern theories now depict creativity as a multifaceted interaction among cognitive processes, environmental conditions, and intrinsic personal characteristics (Amabile, 1983). The widely acknowledged close relationship between creativity and innovation, highlights the idea that previous technical breakthroughs were always produced by breakthroughs in human cognition and creativity (Sarooghi, Libaers, & Burkemper, 2015). In this context, creativity is articulated as the capacity for innovation, defined as the tangible application of inventive thoughts or products (Maravilhas, 2015).

However, the most recent research focused on environmental factors and social contexts, as determinants of creativity, diverge from the argument that creativity's domain is not strictly confined to innovation and discoveries but also includes other prominent and impactful forms of human ingenuity, as expounded in terms of cognitions (Amabile, 1983).

The evolution of the concept of creativity can be traced through various historical periods, each reflecting a distinct understanding of what it means to be creative. J. Dacey (2011) delves into three key historical periods of creativity, highlighting the transition from divine inspiration to contemporary biopsychosocial models, providing a comprehensive overview of how our understanding of creativity has evolved over time (Dacey, 2011). There is a widespread agreement in research that creativity is an attribute that only humans possess. For example, the paper "The evolution of human artistic creativity" by Gillian M. Morriss-Kay describes that the genesis of visual art stands as a hallmark of human uniqueness (Morriss-Kay, 2010). Furthermore, creativity is seen as a trait that is deeply intertwined with human cognitive structures, personality, and emotional intelligence. The findings of "Li et al., 2015" in their study about Brain structure links trait creativity to openness to experience show that creative individuals have higher gray matter volume in their brain (Li et al., 2015). This supports the claim that creativity is a trait unique to humans. Nevertheless, the advent of AI presents fascinating prospects regarding its implications for this quintessentially human attribute, posing questions about the future interplay between human creativity and technological advancement.

2.1 The impact of artificial intelligence on creativity

While AI is typically linked with automation and repetitive activities, recent advances have resulted in the development of sophisticated systems capable of engaging with the creative domain. An example of this is how LLMs, in particular, have demonstrated exceptional performance in tasks such as text and picture generation and analysis (Anantrasirichai & Bull, 2020). The incorporation of AI into the field of creativity ushers in a new age, challenging the contradiction between human creativity and computer competence. Various research have been conducted to investigate the role of AI in enhancing and influencing human creativity, showing a landscape in which computer algorithms not only complement but even improve human creative processes (Chen & Chan, 2023; Boden, 1998). Atkinson and Barker (2023) delve into how AI reshapes creative practices by mediating between human activity and diverse information sources, their analysis suggests that AI's generative capabilities could both compete with human production and foster collaborative creativity, suggesting a collaborative future for AI and human creativity for creative practices (Atkinson & Barker, 2023) Similarly, Barile et al. (2022) investigate the shift from AI as a tool for mechanical tasks to a systemic vision of Intelligence Augmentation (IA), where AI enhances human creativity through the integration of experiences, knowledge, and emotions. This perspective underscores AI's potential to catalyze innovative thinking, bridging gaps between disparate ideas and fostering a rich tapestry of creative outcomes (Barile et al., 2022).

The advent of LLMs represents a significant leap in AI's creative capabilities, extending its influence to text generation, image creation, and beyond. Anantrasirichai and Bull (2020) provide a comprehensive review of AI technologies within creative industries, highlighting AI's role as both a tool and a potential creator their analysis points towards a future where AI, particularly through machine learning algorithms like generative adversarial networks (GANs) and deep reinforcement learning (DRL), becomes an indispensable collaborator in creative activities, augmenting human creativity with its computational efficiency and analytical abilities (Anantrasirichai & Bull, 2020).

Gobet and Sala (2019) emphasize the reciprocal relationship between AI advancements and psychological studies of creativity arguing that AI not only benefits from psychological insights into creative processes, but also provides a rich ground for formulating new experimental designs and theories about human creativity. This synergy between AI and psychology opens up novel avenues for understanding and enhancing the creative capabilities of humans (Gobet & Sala, 2019).

Adding to this discourse, Chen Chan's study, "Large Language Model in Creative Work: The Role of Collaboration Modality and User Expertise," provides a critical examination of how LLMs, such as GPT-3, interact with users in creative tasks. Their research highlights the importance of collaboration modality (how users interact with LLMs) and user expertise, demonstrating that these factors significantly influence the effectiveness of LLMs in creative work. This study underscores the potential for personalized AI tools to enhance creativity, suggesting that the future of creative work will be profoundly shaped by the ways in which humans and AI systems collaborate (Chen & Chan, 2023)

The move of AI from simple job automation to sophisticated creative collaboration represents a significant shift in the creative arena. Boden (1998) presented new ideas, exploratory thinking, and transformational processes, laying the groundwork for contemporary understandings of AI's participation in creativity. (Boden, 1998). More recently, Das and Varshney (2022) investigated the mechanics of generative AI, emphasizing the importance of explicit explanations that foster confidence and ease evaluation of AI-generated creative works, focusing on the importance of openness and interpretability in order to fully realize AI's creative potential (Das & Varshney, 2022).

2.2 The impact of feedback on creativity

Feedback, as a mechanism for delivering information regarding one's performance or output, plays a pivotal role in shaping the trajectory of creative tasks, this influence manifests distinctly across different types of feedback, either being specific or generic, each of which bear unique implications for creativity. Specific feedback, characterized by targeted, detailed insights into one's work, has been shown to significantly enhance the creative process by providing clear guidance and actionable suggestions (Hattie & Timperley, 2007). This type of feedback helps fine-tune and develop creative projects, making it easier for individuals to handle the intricate details of their work with more accuracy. The detailed nature of specific feedback creates a supportive environment that encourages ongoing refinement, which is a key part of the creative process. Conversely, generic feedback, which lacks specificity and tends to be more general in nature, might not offer the same level of direct applicability to creative tasks. However, it can play a crucial role in sustaining motivation and engagement, serving as a form of encouragement that propels individuals to persist in their creative pursuits (Deci, Koestner, & Ryan, 1999). Despite its less tangible impact on the immediate improvement of creative output, the motivational aspect of generic feedback cannot be understated, as motivation is a critical driver of creativity (Amabile, 1996).

The potential of AI-driven feedback in enhancing human creativity emerges as a promising area of exploration, particularly given the capacity of AI to deliver tailored feedback at scale. LLMs, through their ability to analyze vast quantities of data and generate responses based on contextual understanding, are uniquely positioned to provide both specific and generic feedback in a manner that is dynamically aligned with the needs of the individual (Franceschelli & Musolesi, 2023). This tailored feedback mechanism holds the potential to significantly amplify creative outcomes by offering insights that are both relevant and timely.

Furthermore, the theory of the anchoring effect, and cognitive bias elucidated by Kahneman and Tversky (1974), offers additional insight into the potential benefits of AI-driven feedback. The anchoring effect suggests that individuals tend to rely heavily on the first piece of information offered, the "anchor", when making decisions. In the context of creativity, AI-driven feedback can serve as an effective anchor, guiding the direction of creative thought and exploration in a way that can both challenge and expand the creator's original perspective. In that sense, by carefully calibrating the nature of the feedback provided, AI can harness the anchoring effect to foster creative solutions that are both innovative and grounded in relevant context (Tversky & Kahneman, 1974).

To summarize, the influence of feedback on creativity is varied, with various forms of feedback playing unique roles in the creative process. The introduction of AI-driven feedback, underpinned by the concepts of specificity, incentive, and cognitive biases such as the anchoring effect hypothesis, heralds a new era of possibilities for improving human creativity. And as we learn more about AI's ability to provide nuanced feedback, the promise of reaching higher levels of creativity and invention becomes more tangible, signaling a huge advancement in the interplay between human brilliance and AI.

2.3 Research Gap

While the body of literature on the interplay between feedback and creativity is extensive, a deep understanding of how AI-driven feedback specifically influences human creative processes remains unexplored. The growing capabilities of AI, specifically LLMs, in providing context-specific feedback presents a good foundation for research that has yet to be exploited. This section delineates the research gap inherent in the current discourse on creativity, stressing the need for comprehensive studies that further investigate the impact of AI-driven feedback on human creativity.

Existing research has predominantly focused on the general effects of human-generated feedback on creativity, encompassing aspects such as motivation, cognitive development, and the refinement of creative outputs (Zhou, 1998). These studies, while invaluable, do not sufficiently address the unique characteristics and potential of feedback derived from AI systems. The precision, scalability, and adaptability of AI-driven feedback differentiates from traditional forms of feedback, suggesting to have a potentially transformative impact on creative tasks that warrants rigorous examination.

Moreover, the application of cognitive biases, such as the anchoring effect, in the context of AI-driven feedback and creativity introduces another layer of complexity to the dialogue. While Kahneman and Tversky's (1974) work provides a foundational understanding of cognitive biases in decision-making (Tversky & Kahneman, 1974), the implications of these biases within the realm of AI-facilitated creative processes have not been thoroughly investigated. What is more, the potential for AI to both positively and negatively influence creative outcomes through these biases is a critical area of inquiry that remains largely uncharted.

The evolution of AI technologies, particularly LLMs, and their integration into creative workflows also raises questions about the long-term implications of AI-driven feedback on human creativity. Concerns regarding dependency, the deterioration of individual creative skills, and the potential for AI to inadvertently shape creative norms and values, are areas that require thoughtful investigation (He, Shrestha, Puranam, & Miron-Spektor, 2023), the difficult balance between using AI as a tool to enhance creativity and keeping the authenticity and integrity of human creative expression necessitates more study in this area.

2.4 Conclusion

The study of the dynamic interplay between AI and human creativity, particularly via the perspective of AI-driven feedback, has yielded important discoveries and shown key research needs in the field of creativity studies. Key findings from the literature show that feedback has an implicit influence on creative processes, with particular and general feedback playing distinct roles in encouraging innovation. While customized feedback has been shown to improve creative outputs, general feedback is critical for maintaining motivation and interest (Rezwana & Maher, 2022). The emergence of AI, particularly LLMs, provides unprecedented options for tailoring feedback, possibly enhancing human creativity in new ways. The significance of our research question, investigating the extent to which feedback from LLMs influences the creativity and quality of human work, gains profound relevance in this context. As the literature reveals, while the effects of human-generated feedback on creativity have been extensively studied, the specific influence of AI-driven feedback remains a largely untapped area. This gap underscores a pivotal opportunity to delve deeper into the mechanisms through which AI can enhance or hinder creative processes, an inquiry that is not only timely but essential in the age of rapid technological advancement. Our research, therefore, seeks to bridge this identified gap by employing a methodology that directly examines the impact of LLM feedback on creative outcomes. Through empirical analysis, we aim to delineate the characteristics of effective AI-driven feedback and its psychological effects on creators, thereby contributing a nuanced understanding to the field. This approach not only addresses the gap highlighted in the literature but also paves the way for future research endeavors to further elucidate the complex relationship between AI and human creativity. In conclusion, this literature review has laid the groundwork for a deeper investigation into the role of AI-driven feedback in creative contexts, highlighting its potential benefits and pitfalls. By situating our research question within the broader discourse on creativity and AI, and by outlining a methodology designed to delve into this intersection, our study stands to contribute valuable insights into the evolving landscape of creativity in the digital age. As we move forward, the findings from this research hold the promise of bring forward theoretical frameworks and practical applications, enriching our collective understanding of creativity in the era of AI.

3 Theoretical Framework

3.1 Creativity

According to the definition of "Sternberg and Lubart" (1999) creativity refers to the ability to generate novel (original, unexpected) and appropriate (useful, effective) ideas, products, or solutions (Sternberg & Lubart, 1998). It's a complex process influenced by both individual capabilities (e.g., knowledge, skills) and environmental factors (e.g., feedback, motivation). This multidimensional understanding highlights creativity as a dynamic interplay within the creative process (Amabile, 2012). Crucial to this process are two distinct thinking modes: divergent thinking, where individuals explore a wide range of possibilities to generate a multitude of ideas, and convergent thinking, where they evaluate and refine these ideas to select the best solution (Cropley, 2006). Ultimately, intrinsic motivation, the internal drive to engage in a creative task, is a key factor in fostering creativity as it fuels the creative journey by encouraging exploration and perseverance. When individuals are intrinsically motivated in a creative task, they are more likely to persevere through challenges and explore new ideas (Ryan & Deci, 2000). These core elements of creative processes are essential for examining how LLM feedback, and its variation, might influence human creativity. Within the scope of this research, understanding creativity's multidimensional character is crucial, especially in evaluating its role in creative processes facilitated by AI. The emphasis on individual and environmental influences connects directly to the theories on feedback mechanisms and human-computer interaction (HCI). By examining creativity through this comprehensive lens, we establish a solid foundation for investigating how AI-driven feedback can interact with and potentially amplify human creative capabilities.

3.2 Feedback mechanisms

Effective feedback plays a significant role on the creative process, guiding individuals through this process and helping them refine their work. More precisely it is crucial for identifying weaknesses, and ultimately transition from divergent exploration to convergent selection (Hattie & Timperley, 2007). The research of Hattie et. al. found out that the impact of feedback can either be positive or negative on learning and achievement depending on the type and delivery of the feedback. This becomes critical to our research when exploring how variations in LLM feedback influence key stages of creative thinking, from divergent exploration to convergent selection, and in designing AI-driven feedback mechanisms to optimally support creative processes. It is essential to consider the effectiveness of these systems in providing various types of feedback and any potential limitations for specific categories, allowing for a deeper investigation into the nuanced impact of LLM feedback on creative thinking stages. Furthermore understanding how different feedback types influence intrinsic motivation and the identification of weaknesses is crucial for assessing the overall impact of LLM feedback on creative output. Specific, constructive feedback not only enhances motivation by offering clear improvement directions and recognizing effort but also identifies areas needing attention, thereby facilitating the creative process. On the other hand, vague or overly critical feedback can diminish motivation, lacking actionable guidance and focusing disproportionately on deficiencies without highlighting strengths (Hattie & Timperley, 2007). This knowledge plays a vital role in interpreting the results of this study as motivation might influence creative results. These insights underline feedback's complexity and its enhancement potential through LLMs, from ideation to refinement. Effective feedback, crucial for transitioning from divergent exploration to convergent selection, relies on understanding feedback variations (Hattie & Timperley, 2007).

3.3 Human-Computer Interaction (HCI)

Human-Computer Interaction (HCI) is the study of making computer systems accessible and effective by understanding and improving how people interact with computers, with an emphasis on user-friendly interfaces and individualized interactions. The HCI discipline is critical for understanding how people engage with AI systems like LLMs, and how this contact affects the feedback process in creative work (Fischer, 2001). This allows us to better understand this dynamic and optimize creative outputs. G. Fischer (2001) highlights the importance of making systems more usable and useful by tailoring them to fit users' specific background knowledge and objectives. This approach underscores the necessity of user modeling in HCI, aiming to personalize the interaction between humans and AI systems, including LLMs, to improve the creative feedback process (Fischer, 2001)

Several studies argue for a deeper integration between humans and computers, shifting from traditional human-computer interaction to more seamless human-computer integration, where both entities act autonomously, suggesting a more holistic approach to behavior patterns. This perspective, advocated by Umer Farooq (2016), is especially relevant for understanding how LLMs can serve as integral partners in creative processes, enhancing the symbiosis in feedback mechanisms (Farooq & Grudin, 2016). Similarly, Ben Shneiderman's Eight Golden Rules of Interface Design (Yusof, Amin, Zainudin, & Baker, 2004) emphasize making interfaces user-friendly and intuitive, which is crucial for fostering efficient and satisfying interactions. By integrating design principles like consistency, informative feedback, and error prevention into AI systems, especially in creative environments, we ensure that interfaces do more than just support—they actively enhance the creative process. This makes interactions intuitive and feedback both constructive and meaningful. It highlights the significance of thoughtfully designing survey tasks and carefully crafting LLM prompts to optimize the user experience. Incorporating feedback specificity into HCI design principles is particularly vital for enhancing LLM feedback systems in creative tasks. Specific, actionable feedback transcends generic praise or criticism, offering concrete suggestions for improvement and facilitating a more productive and engaging creative process (Hattie & Timperley, 2007).This technique fits with user-centered design by directly addressing the user's work and helping them to better outputs, indicating a vital integration of HCI principles with the usefulness of LLMs in promoting human creativity.

3.4 AI and Creativity

The intersection of AI and human creativity is a rapidly growing field of study that poses basic concerns about the nature and boundaries of both human and machine creativity. In the following we will discuss significant research findings to investigate how AI technology might assist rather than replace human creativity. According to Das and Varshney (2022) explaining the workings of generative/creative AI algorithms—or the artifacts they produce—can enhance trust, enable action, and provide a basis for evaluation. This transparency is crucial for integrating AI into creative processes, ensuring that AI serves as an augmentative tool rather than a replacement for human creativity (Das & Varshney, 2022) Furthermore the study "Large Language Model in Creative Work: The Role of Collaboration Modality and User Expertise" by Chen and Chan explored how AI aids, rather than replaces, human creativity. It investigates two collaboration methods: using AI as a "ghostwriter" (leading content generation) and a "sounding board" (providing feedback). Results show that the sounding board method enhances ad quality, particularly for non-experts, by helping them achieve content closer to expert-level quality. This suggests AI can augment human creativity by offering ideas and refining concepts, but doesn't replace the human touch in crafting truly successful content. Humans working collaboratively with LLMs (feedback mode) rather than employing the LLM as a ghostwriter considerably improves creativity and job quality. This demonstrates that there is no alternative in creative job situations.(Chen & Chan, 2023). Lastly, Anantrasirichai and Bull (2020) review AI's applications within the creative industries, differentiating between AI as a creative tool and its potential as a creator. They foresee AI being widely adopted as a collaborative assistant, enhancing rather than replacing human creativity. This perspective underscores the human-centric potential of AI in creative industries, emphasizing augmentation over automation, which becomes vital for our understanding of how to further enhance this collaboration in the creative realm (Anantrasirichai & Bull, 2020). According to Gobet and Sala (2019), significant developments in AI have not only challenged our understanding of human rationality but also revealed the creative potential of machines. Their work posits that AI can design new classes of experiments in psychology, promising more profound insights into creativity, thus opening new avenues for understanding human creativity through the lens of AI (Gobet & Sala, 2019). Further, a study by Norton, Heath, and Ventura (2013) demonstrate the potential of AI to engage in creative acts through their development of DARCI, a system designed for image creation. By evaluating DARCI's performance based on appreciation, imagination, and skill, they assert that artificial systems can indeed exhibit creativity, thus supporting the idea that AI can complement human creativity (Norton et al., 2013). Lastly, Boden's (1998) distinction between computational creativity (generating creative outputs) and psychological creativity (human novelty and understanding) highlights AI's potential to support the former and its limitations in replicating the latter. This distinction is vital for recognizing the complementary roles of AI and human creativity, suggesting that while AI can generate novel combinations and explore conceptual spaces, it lacks the depth of human understanding and context (Boden, 1998).

3.5 AI and Machine Learning

LLMs leverage AI and machine learning (ML) techniques to analyze information and provide feedback (Zhao et al., 2023). Understanding these underlying principles is essential for appreciating how LLMs operate within creative environments. ML algorithms form the backbone of most AI-driven creative systems. ML involves computer algorithms that enhance their performance through experience. This adaptive nature allows AI systems to offer personalized feedback in creative tasks by analyzing extensive data to discern patterns and relationships. For instance, in the realm of LLMs, this might involve processing text, code, or various creative content forms. Through this analysis, LLMs develop the capability to identify effective creative strategies and pinpoint areas needing enhancement, thereby delivering feedback specifically tailored to the user's task. An example could be an LLM reviewing a draft of advertising copy and suggesting stylistic refinements or identifying parts that lack clarity (Jordan & Mitchell, 2015).

Despite their proficiency in pattern recognition and content generation, it's crucial to recognize the limitations of current ML techniques in handling creative tasks that require nuanced understanding, emotional intelligence, and a broader contextual grasp. In such scenarios, human judgment remains indispensable for fruitful creative collaboration.

3.6 Assessment Tools to Evaluate Creativity and Quality

To assess semantic creativity and adherence to the creative task instructions, it becomes important to understand how we employed semantic similarity analysis, which measures how closely the generated text aligns with a reference style or prompt. In this study, SentenceTransformer (SBERT), a pre-trained neural network model, was used to compute sentence embeddings—numerical representations that capture the semantic meaning of a sentence. By comparing the embeddings of the generated text with the reference text, we obtained a similarity score, with higher scores indicating closer alignment to the intended creative direction, potentially reflecting a stronger grasp of the creative task and its goals. Similar Natural Language Processing (NLP) tools like Spacy and LABSE were also used alongside SBERT for a comprehensive semantic similarity analysis which analyze semantic similarities similarly (Reimers & Gurevych, 2019). Building on semantic considerations, we also explored the emotional dimension of creativity using sentiment analysis with Roberta, a pre-trained deep learning model for natural language processing, to analyze the emotional tone of the generated text. This focus on emotional depth, examining the range and complexity of emotions expressed, complements the semantic analysis, as engaging with a wider range of emotions can foster creative thinking (Camacho-collados et al., 2022; Loureiro, Barbieri, Neves, Espinosa Anke, & Camacho-collados, 2022).

Our analysis further delves into the semantic quality of the generated text by employing readability metrics such as the Flesch-Kincaid Grade Level and the Coleman-Liau Index. These tools are instrumental in measuring text complexity and ease of reading, which are vital for assessing how creative texts might diverge into more complex or unconventional structures while still maintaining clear communication. The Flesch-Kincaid Grade Level is designed to reflect the U.S. school grade level needed to understand the text. It calculates this based on the number of syllables per word and words per sentence; a higher grade level suggests more complex language that might be harder for wider audiences to understand. Similarly, the Coleman-Liau Index predicts the grade level required to comprehend the text based on the number of characters per word and sentences per 100 words, thus focusing more on character count rather than syllables. This index is particularly useful in indicating how dense or accessible the text is. By using these measurements, we get insights into the semantic quality of creative outputs, presenting a balanced picture of how inventive expressions impact readability and, as a result, the text's accessibility to its intended audience. This dual approach not only shows the complexities of creative writings, but also emphasizes the significance of developing messages that remain intelligible, ensuring they resonate well with their audience (Webster, 2019; Choudhery et al., 2020). Building upon the quantitative and semantic metrics, our study further employed the Roberta model for sentiment analysis to evaluate the emotional dimensions of the generated texts. This analysis sought to uncover the diversity and range of emotions expressed in the creative outputs, aligning with theories suggesting that more varied sentiments indicate higher creativity. According to Lubart and Getz (1997), shifts and surprises in narrative that evoke multiple emotions require a deeper level of emotional thinking compared to monotonous storylines that generate fewer emotional responses (Lubart & Getz, 1997). Similarly, Gutbezahl and Averill (1996) highlight that individual differences in emotional creativity manifest not only in the complexity but also in the variety of emotions portrayed in both textual and visual expressions (Gutbezahl & Averill, 1996). By integrating sentiment analysis with Roberta, we were able to quantitatively assess the emotional breadth within the texts, correlating diverse emotional expressions with enhanced creative capacity. This nuanced examination of sentiments provides a more comprehensive understanding of how effectively the participants engaged with the creative tasks, further enriching the interpretative depth of our semantic and readability analyses.

4 Hypotheses

In this research, we explore the extent to which feedback provided by Large Language Models (LLMs) influences human creativity, addressing a gap in understanding the interactive dynamics between artificial intelligence and creative processes. The formulation of our hypotheses is grounded in both theoretical perspectives on creativity and empirical evidence suggesting that feedback can significantly impact creative output. By systematically investigating various feedback modalities—ranging from specific to general, and even the absence of feedback—we aim to elucidate how these different approaches affect the quality and creativity of human work. This study not only seeks to identify the potential of LLMs as tools for enhancing creativity but also to understand the conditions under which their influence is most effective. The hypotheses outlined below are designed to test these interactions comprehensively, providing insights that could guide future applications of AI in creative domains.

4.1 Specificity of Feedback and Creativity

The first hypothesis explores the correlation between the specificity of feedback from Large Language Models (LLMs) and the resulting creativity of human outputs. The Alternative Hypothesis (H1a) posits that specific, detailed feedback provided by LLMs, positively influences the creativity of human work. It is expected that individuals who receive detailed guidance will produce more creative outputs than those given generic feedback. Conversely, the Null Hypothesis (H0a) suggests that there is no significant difference in creativity between individuals receiving specific versus generic feedback, indicating that the level of detail in the feedback may not impact creative outcomes.

4.2 Feedback on Different Creative Tasks

The second hypothesis addresses the variability in the effectiveness of feedback depending on the nature of the creative task. The Alternative Hypothesis (H1b) asserts that the impact of feedback specificity from LLMs varies with the type of creative task, being more beneficial for tasks demanding high originality than for tasks focused on persuasive communication. The Null Hypothesis (H0b) challenges this, proposing that the influence of feedback specificity does not differ significantly across various types of creative tasks used in this study.

4.3 Specificity of Feedback and Creativity

This hypothesis examines whether the presence of any feedback (specific or general) enhances the creativity of outputs compared to no feedback at all. The Alternative Hypothesis (H1c) suggests that receiving feedback, regardless of its specificity, results in higher creativity in human outputs. The Null Hypothesis (H0c) counters this by hypothesizing that there is no significant difference in creativity between participants who receive feedback and those who do not, implying that feedback might not inherently enhance creativity.

4.4 Sentiment Depth on Different Modalities

In the fourth hypothesis, the focus shifts to the emotional depth of the creative outputs under different feedback conditions. The Null Hypothesis (H0d) assumes that there is no significant difference in the average sentiment depth between specific and general feedback conditions, while the Alternative Hypothesis (H1d) posits that creative outputs from the specific feedback condition exhibit greater sentiment depth than those from the general feedback condition. This hypothesis aims to explore how the nature of feedback affects the emotional resonance of creative texts.

4.5 Readability of Feedback and Text Quality

Finally, the fifth hypothesis evaluates whether the complexity of feedback affects the readability and overall quality of the creative texts produced. The Alternative Hypothesis (H1e) posits that feedback provided in the specific condition is more complex, as indicated by lower readability levels (higher Flesch-Kincaid Grade and Coleman-Liau Index scores). In contrast, the Null Hypothesis (H0e) suggests that there is no significant difference in the readability levels of feedback between the specific and general conditions, challenging the assumption that more specific feedback is inherently more complex

5 Methodology

This section lays out the methodologies used in the research for analyzing the influence of feedback from LLMs on the originality, creativity, and quality of human-generated work. The cornerstone of this study is the hypothesis that LLMs can significantly augment human creativity through targeted feedback. According to our foundational theories, innovation is a prime indicator of creativity, as it directly relates to the novelty and uniqueness of outputs, a critical dimension in the creative process (Reimers & Gurevych, 2019). This idea was selected to align with the theoretical assertion that creativity can be effectively measured by the level of innovation of the work produced, seeking for relevant and robust measurements (Anderson, Potočnik, & Zhou, 2014). In addition to analyzing the direct impacts of LLM feedback on creativity, the design of the research aimed to mitigate potential anchoring effects that could lead respondents to fixate on initial ideas presented via feedback, influencing their creative outputs (Tversky & Kahneman, 1974). By systematically varying the specificity and type of feedback provided by LLMs, we aimed to explore how different feedback modalities influence the creative process and whether initial feedback sets a cognitive benchmark that affects further idea development. Especially that is why the respondents were instructed to initially write their text on their own and then obtaining certain LLM feedback which was then based on the already written ideas. The methodology encompasses several key components: the selection and design of creative tasks, the operationalization of feedback specificity, the process of capturing and analyzing creative outputs, and the statistical methods used to assess the effects of LLM feedback on creativity and the semantic quality.

Participants engaged in two creative tasks designed to challenge their creativity: a Story Continuation Task, where they wrote the continuation of a story about a magical ship navigating the Caribbean seas, and an Advertisement Task, where they created a compelling advertisement for a computer mouse. For these tasks, the specific feedback group received detailed feedback focused on narrative engagement and style, while the general feedback group received broader, less detailed feedback. Participants were instructed to submit their initial texts to ChatGPT via a pre-prompt underneath the task that clearly outlined the type of feedback they should expect, encouraging them to incorporate this feedback into at least one subsequent revision of their text. The control group, receiving no feedback, provided a baseline for comparison. The use of the Qualtrics platform was instrumental in distributing tasks, collecting responses, and ensuring uniform implementation of feedback modalities across participant groups, thereby enhancing the study's ability to measure the relative effects of feedback specificity on participant outcomes. This approach not only standardized the data collection procedure but also improved the data quality, revealing important insights into the dynamics of AI-enhanced creativity.

This method recognizes possible limits, particularly when using innovation and emotional depth as the only measures of creativity. Although innovation gives a quantitative measure of originality, creativity is a complicated construct impacted by a variety of characteristics based on more than one feature. Thus, while focusing on invention provides for a more specific study, further research into other factors as a measure of creativity is required.

5.1 Research Design

This study examines how different types of feedback from an LLM influence human creativity, through a between-subjects experiment. This design was chosen for its effectiveness in isolating the study's variable of interest—feedback specificity—, being able to measure its distinct impact on the creativity, and quality, of content across diverse participant groups. Moreover, this experimental approach helped mitigate the risk of an anchoring effect, where respondents might become anchored to an initial idea of the feedback, potentially influencing subsequent responses to feedback provided by the LLM and viewed by the participant (Tversky & Kahneman, 1974). This was a major reason why the two creativity tasks are completely distinct in topics. To assess the effectiveness of feedback from the large language model ChatGPT, three feedback modalities were defined. The first modality, highly specific feedback, provided participants with detailed, specific feedback, giving the text a clear direction and provided the author with more precise ideas on how to revise the text. The second modality, general feedback, involved offering broader, less detailed feedback, being less suggestive on the development of the text. Lastly, the control group, received no feedback, providing a baseline for comparison, with the previous two cases. Respondents who were not part of the control group were given a pre-defined prompt beneath the task instruction, which they had to copy and paste to ChatGPT along with their assignment in order to obtain feedback. The use of the Qualtrics platform was instrumental in distributing tasks, collecting responses, and ensuring uniform implementation of feedback modalities across participant groups, thereby enhancing the study's ability to measure the relative effects of feedback specificity on participant outcomes.

5.2 Participants

Participants were recruited from a broad demographic spectrum to encompass a wide range of creativity and content generation skills. Recruitment channels included social media platforms, university groups, and direct contacting, aiming for a diverse participant pool. In order to broaden the cultural mix of participants, in addition to Germans, a significant number of people from Ecuador, South America, responded in Spanish. To ensure an unbiased evaluation of feedback modalities, each participant was allocated randomly to one of the three research groups. Following an initial gathering of demographic information, participants proceeded with the creative tasks, which were uniform across each feedback modality but varied in the specificity and detail of the feedback provided by ChatGPT. These tasks included writing a creative story continuation from a given first sentence and crafting a compelling advertisement for a computer mouse, where no initial sentence was provided. The only distinction among the modalities was the specific prompt that participants, except those in the control group, were required to copy and paste into ChatGPT. This setup allowed the experiment to investigate the impact of feedback quality on the creative process and its outcomes comprehensively.

5.3 Measures of Creativity and Quality

The evaluation of creativity and quality in the participant texts used a mixed-methods approach, integrating different quantitative analyses to provide a more comprehensive assessment of the results. For a quantitative analysis of results, creative texts were subjected to similarity scoring using SBERT and SpaCy models through the Python SentenceTransformer library. Lower scores in this context indicated higher novelty, suggesting greater originality in the content produced under different feedback conditions. Creativity was directly assessed by evaluating semantic similarities using two SBERT models and indirectly measured by utilizing the Roberta model to assess creativity characteristics based on emotional depth of the sentences. It was expected that participants engaged in creative writing tasks that challenged their originality, that is why the tasks were designed with clear objectives and constraints that required text generation under each feedback condition. Creativity of the task could be assessed on the criteria of novelty, relevance, and diversity. This approach ensured that the assessment aligns with a well-supported research on creativity evaluation, providing a solid basis for interpreting the impact of feedback on creative outputs. Similarly, semantic quality of the answers was also assessed through innovation, and various readability metrics, including the Flesch-Kincaid Grade Level and Coleman-Liau

Index, which estimate the readability based on sentence length and word complexity. Additional assessments covered grammar and mechanics through Language Model assessment, along with coherence and cohesion evaluations, providing a multidimensional view of the text quality.

5.4 Data Collection Methods

We used Qualtrics, a robust online survey platform, to facilitate both the collection of demographic information, and the submission of creative tasks by participants. This platform was chosen for its reliability and ease of use, which are essential for maintaining consistency in data collection across a diverse participant group. The distribution of the survey was strategically executed through social media, university groups, and direct contacting to ensure a wide reach. Specifically, we utilized student WhatsApp groups, predominantly from the University of Cologne, to target participants who are likely familiar with technological tools such as ChatGPT. This approach not only facilitated the recruitment of participants who are generally well-versed in using advanced language models but also increased the likelihood of receiving well-articulated and thoughtful creative submissions, given academic rank of the university. Additionally, social media platforms like Instagram were employed to further broaden the participant's pool beyond university setting. Participants in the survey were provided with clear instructions on how to proceed with their submissions. For those that were not part of the control group, specific guidance was given on how to interact with ChatGPT. They were instructed to copy a provided prompt from beneath the task description into the ChatGPT interface to receive feedback. Afterwards, participants were required to revise their initial text at least once, based on the feedback received. This procedure was designed to mimic real-world conditions where creative feedback is often iterative, and refining ideas based on initial feedback is a critical part of the creative process. The survey's established rules guaranteed that all participants, regardless of past experience with AI-driven feedback systems, could effectively interact with the job. This methodological approach not only standardized the data gathering procedure, but it also improved the data quality, revealing important insights into the dynamics of AI-enhanced creativity.

6 Experimental Setup

A systematic and programmatic approach was adopted to collect and analyze data using Python, ensuring precision in managing and interpreting the diverse range of variables captured during the experimental process. Comprehensive descriptive statistics were then calculated for each feedback modality—specific, general, and control (no feedback) to provide an initial overview of the data collected from participants. These statistics were crucial for summarizing demographics and the central tendencies and variability of the creativity scores and other relevant metrics derived from participant submissions.

A critical measure tracked during the survey was the time taken to complete it, to ensure data integrity. The calculations of the total and average time spent on tasks, the standard deviation of time spent, the total and average length of the story continuations, and the standard deviation of these continuations, provided insights into the engagement levels and effort exerted by participants across different modalities, offering a preliminary view of the impact of feedback specificity. Outliers in this variable were addressed programmatically, treating any time recording exceeding 100 minutes as outliers, that were replaced with the mean time of completion for that particular feedback modality, thereby normalizing the data for more accurate analysis. Additionally, for linguistic diversity, the same statistics were once again separately calculated for German, English, and Spanish speakers; including the average time to complete the survey, average age, and the total number of participants per language group. This segmentation allowed for assessing whether language influenced the responsiveness to LLM feedback and the creativity displayed in the tasks. Collecting and analyzing these metrics systematically allowed the research to adhere to rigorous scientific standards and ensured that the data reflected the real-world complexities of human creativity as influenced by AI-driven feedback. This methodological thoroughness provided a robust foundation for the subsequent stages of data analysis and interpretation, setting the stage for insightful conclusions about the interplay between AI feedback and human creative output.

For the analysis of the impact of LLM feedback on the creativity of participant submissions across different modalities and languages, we used advanced natural language processing (NLP) tools and frameworks to quantify the uniqueness and novelty of participant responses to creative tasks. The analysis began with the utilization of the SpaCy framework to calculate the average textual similarity for each modality group, encompassing both the story continuation and the advertisement tasks. This provided a baseline measure of how much participant responses deviated from each other within the same feedback conditions, indicating levels of textual originality.

Moreover, the study expanded its semantic analysis capabilities by incorporating the SentenceTransformers library, specifically the multilingual 'paraphrase-multilingual-MiniLM-L12-v2' model. This multilingual SBERT (Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks (Reimers & Gurevych, 2019)) model enabled the calculation of average textual similarities for each modality across the given tasks, supporting a more distinctive understanding of creativity in multilingual contexts. To enhance the precision of semantic similarity measures, all responses from participants, originally in German and Spanish, were translated into English, allowing the application of the 'all-MiniLM-L6-v2' model from Sentence Transformers, which is specifically trained on English data and therefore provides more precise results than its multilingual counterparts (Reimers & Gurevych, 2019). However, it is important to note the potential limitations of translation of the texts, as it may introduce some loss of creative characteristics, inherent in the original language submissions.

In conjunction with semantic similarity, sentiment analysis was conducted using the 'cardiffnlp/twitter-roberta-base-sentiment-latest' model, which is tailored to process English texts. This analysis was applied to the translated submissions, calculating mean sentiment scores for each task and modality; these sentiment scores helped to assess the emotional depth and resonance of the creative outputs, with metrics segmented into negative, neutral, and positive categories (Barbieri, Camacho-Collados, Neves, & Espinosa-Anke, 2020). Additionally, a similar sentiment analysis was performed using the 'siebert/sentiment-roberta-large-english' model to confirm and compare the sentiment findings, ensuring robustness and consistency in the emotional analysis of the participants' texts (Camacho-collados et al., 2022)

We used two sentiment analysis models because they have fundamental distinctions that compliment each other exceptionally well when it comes to assessing semantic sentiments. Cardiff's model is a basic RoBERTa, which may be faster for inference but less sophisticated in sentiment detection. In contrast, Siebert's approach makes use of a big RoBERTa, which has a better capability for capturing complicated emotion patterns but may require more computer resources. Employing both models can result in a more comprehensive evaluation. The base model (Cardiff) provides a rapid first sentiment analysis, whilst the big model (siebert) allows for a more in-depth study of possibly subtle emotional complexities within text data. This complementary strategy can improve dependability and even disclose features of emotion that a single model may overlook (Loureiro et al., 2022; Hartmann, Heitmann, Siebert, & Schamp, 2023).

7 Results

We present our findings organized into several distinct segments, each addressing different aspects of creativity as influenced by varying degrees of feedback provided by the Large Language Model ChatGPT. This segmentation follows the theoretical framework outlined for this research, mainly emphasizing the multifaceted nature of creativity.

7.1 Descriptive Statistics

Our study was designed to assess the impact of specific, general, and no feedback on the creativity of human outputs across various tasks. We were able to identify underlying trends and tendencies, by analyzing the descriptive statistics of our collected data.

The survey sent through different channels, received 99 responses, from which only 33 were considered as valid, after excluding entries where the creative tasks were not completed. Excluded responses typically spent an average of just two minutes on the survey, filling out only the demographic information. After this preliminary screening, further data cleansing was conducted, particularly addressing outliers in survey completion times; specifically, any response time exceeding 100 minutes was adjusted to reflect the mean time spent, that participants of that feedback modality needed to complete the survey.

In this initial section of our results, we analyze the data collected across the three feedback modalities, to uncover trends and patterns in time investment, submission length, and participant engagement. Participants in the specific feedback group, totaling eight individuals, spent an aggregate of 336 minutes on tasks, averaging 42 minutes per session with a standard deviation of 22.42 minutes. This group contributed 6,830 words for the story continuation task and 3,057 words for the advertisement task, averaging 853.75 and 382.13 words respectively, demonstrating significant engagement and creative output.

In the general feedback group, made up of seven participants, a total of 183.17 minutes to fill in the survey was recorded, averaging 26.17 minutes per session. The story outputs from this group totaled 6,841 words, indicating a higher average word count per participant at 977.29, although with a greater variance as reflected by a standard deviation of 467.45 words. The advertisement submissions mirrored this trend with a total of 4,408 words, averaging 629.71 words per participant. In contrast, the control group, which included 18 participants receiving no AI feedback, spent a total of 272.12 minutes on the tasks. They averaged significantly less time per session at 15.12 minutes and produced 8,586 words for story continuation and 6,806 words for advertisement tasks, averaging 477 and

378.11 words respectively. This highlights a potentially lower level of engagement or efficiency in task completion without the directional influence of feedback.

Expanding our analysis to include language-specific demographics, the German participants (N=21) averaged 29.68 minutes to complete the tasks, with a collective time investment of 623.28 minutes, indicating a higher level of engagement. The average age of this group was 26.07 years, distributed among 11 females and 10 males. In comparison, the English-speaking participants (N=4) completed their tasks in an average time of 16 minutes and had an average age of 29.88 years. The Spanish speakers (N=8) spent even less time on average to complete the tasks, with 13 minutes, while their average age is of 24.5 years. These results show varied engagement across linguistic groups, which may reflect cultural differences when approaching the tasks, or the level of familiarity with the technology used.

This first descriptive approach on the results highlights the differences in engagement and output across various feedback modalities but also illustrates how demographic factors such as language can influence the interaction with the experimental setup. This provides an early understanding of the sample for subsequent inferential statistical analyses aimed at assessing the impact of LLM feedback on creativity and quality in human texts.

Table 1: Summary Statistics

	No AI (N=18)	Specific Feedback (N=8)	General Feedback (N=7)
	Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)
Age	30 (5)	32 (6)	31 (5)
Task Duration (mins)	15.12 (10.36)	42.00 (22.42)	26.17 (9.69)
Length (Story)	477 (290.53)	853.75 (419.44)	977.29 (467.45)
Length (Ad)	378.11 (162.58)	382.13 (136.79)	629.71 (317.25)
Gender			
	N Pct	N Pct	N Pct
Female	10 56%	4 50%	4 57%
Male	8 44%	4 50%	3 43%
Education			
	N Pct	N Pct	N Pct
Abitur	4 22%	2 25%	2 29%
Bachelor	9 50%	4 50%	3 43%
Master	4 22%	2 25%	2 29%
Other	1 6%	0 0%	0 0%

7.2 Semantic Similarity Analysis

Semantic similarities refer to the degree to which different texts share meaning or conceptual content, and they are crucial in creativity as they indicate how closely ideas or expressions align with established creative or innovative standard (Reimers & Gurevych, 2019). Analyzing these similarities is essential in understanding how different types of feedback can influence the content's alignment with certain creative benchmarks. The primary aim is to discern whether specific, general, or no feedback affects the semantic proximity of participant responses to predefined creative standards. The results from the semantic similarity analysis using spaCy provide a foundational understanding of how linguistic feedback impacts the alignment of participant responses with established creative benchmarks. Initially, the analysis utilized an English large NLP model from spaCy to evaluate the semantic content of responses. It is important to note that the initial results might exhibit a bias, given the model's primary training on English language data, which could affect the precision of the semantic similarity scores for responses in German and Spanish. To address potential biases and enhance the validity of the findings, participant responses were translated into English with the state-of-the art translator "DeepL". Furthermore, for a subset of the data where German was predominantly used, a German language model was employed, providing a more accurate assessment for those specific instances as it is original data. Here, the English model results showed that the specific feedback modality group exhibited higher average similarity scores both in the creative story continuation task (0.751) and the advertisement copywriting task (0.674) compared to the general feedback (0.582 for story continuation, 0.547 for advertisement) and control groups (0.486 for story continuation, 0.472 for advertisement). This suggests a higher semantic alignment with the desired creative standards when specific feedback is provided.

Additionally, the results using the German model, where the sample size was largest, and thus potentially more reflective of the modality's effect, indicate interesting variances. The specific feedback modality yielded lower average similarity scores (0.821 for story continuations and 0.744 for ads) compared to the general feedback (0.867 for story continuations and 0.844 for ads) and control groups (0.835 for story continuations, 0.822 for ads). This is an intriguing divergence from the results observed with the English translations, possibly reflecting differences in how feedback is processed across languages or the nuances captured by the different models.

Following the spaCy analysis, the Sentence Transformer framework (SBERT) was applied to further assess the deep semantic relationships within the texts.

SBERT, known for its effectiveness in generating embeddings that capture underlying semantic meanings, provided a complementary perspective to the structural insights offered by spaCy. This phase of the analysis aimed to discern not just the overt textual similarities but also the subtler semantic connections that might be influenced by feedback specificity. The use of multiple models for analysis provides a more refined understanding of the intricate semantic relationships present in the participants' responses, given their capacity to handle multiple languages and capture deeper semantic meanings. The multilingual capabilities of the SBERT model, particularly the 'sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2', were instrumental in analyzing texts in their original languages. This approach ensured that the semantic integrity of the responses was maintained, avoiding potential distortions from translation. The results from this model showed that the general feedback modality led to slightly higher average similarity scores in story continuations (0.5434) compared to the specific modality (0.5188), and even higher in the ad computer mouse task (0.5867 for general vs. 0.4397 for specific). Interestingly, the control group's scores were comparable to those of the general feedback in the ad task (0.5837), suggesting that the absence of directed feedback does not necessarily diminish semantic alignment in more structured tasks.

Parallelly, the LABSE model, another semantic similarity model developed by the 'SentenceTransformer' framework, offered additional insights with its robust multilingual processing capabilities. The scores from LABSE highlighted a consistent trend across tasks, with the general modality often exhibiting comparable or slightly lower semantic similarity scores than the specific modality in the story continuation task (0.5574 for general vs. 0.5667 for specific) but not in the ad task. This suggests that the nature of the task may influence how feedback affects semantic alignment.

Further refining the analysis, the SBERT all-MiniLM-L6-v2 model, was employed after translating all non-English assignments into English. This model, trained extensively on English data, aimed to provide a high-resolution view of semantic similarities but also introduced a risk of semantic loss due to translation. Here, the specific modality displayed varied performance, with lower similarity in story continuations (0.4424) but higher in ad tasks (0.5767) compared to the general and control groups. By comparing results from various NLP tools and accounting for translation effects and model specificity, we gain a nuanced understanding of how feedback influences creative expression. These initial findings lay the groundwork for further discussions that will connect these patterns to theoretical frameworks and related studies, exploring their implications for improving creative processes with targeted feedback.

These findings facilitated the evaluation of the research hypotheses, addressing the central research question. To assess the hypotheses that address the semantic similarities, statistical methods, including t-tests and a two-way ANOVA, were employed to examine the impact of feedback on creative tasks.

The t-test comparing semantic similarity scores between specific and general feedback modalities, as measured using both the multilingual (sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2) and English (all-MiniLM-L6-v2) SBERT models, both revealed statistically significant differences (p-values < 0.05). These results indicate a greater semantic convergence towards established creative benchmarks in the specific feedback group compared to the general feedback group. Notably, the English data, analyzed with the more precise all-MiniLM-L6-v2 model, supported these findings, suggesting that the specificity of feedback significantly influences creative outputs. Results from the story continuation task showed no significant difference in creativity scores between specific and general feedback ($t(78.39) = -0.607$, $p = 0.546$). Conversely, the advertising task demonstrated a substantial increase in creativity with specific feedback compared to general feedback ($t(70.01) = -4.219$, $p < 0.001$). An overall analysis across both tasks also revealed that specific feedback led to significantly higher creativity than general feedback ($t(193.99) = -3.185$, $p = 0.002$), supporting the claim that feedback specificity positively impacts creative outcomes. These results support the first hypothesis, concluding that the specificity of feedback from LLMs is positively correlated with the creativity of human work.

The t-test comparing specific feedback to general feedback across tasks indicated a significant difference with a p-value lower than 0.05, leading to the rejection of the null hypothesis (H0a). This suggests a statistically significant impact of feedback specificity on the creativity of human work. Additionally, separate t-tests for each task type—story continuation and advertisement copywriting—yielded mixed results. The t-test for story continuation did not show a significant difference (p-value higher than 0.05), while the t-test for advertisement copy showed a significant difference, with a p-value lower than 0.05. Further analysis using the English dataset with the all-MiniLM-L6-v2 model also resulted in significant differences with a p-value lower than 0.05, thereby rejecting the null hypothesis for Hypothesis 1 once more, confirming a significant impact of feedback specificity.

The two-way ANOVA conducted to assess the interaction between feedback type and task type yielded results that also impacted hypothesis testing. The interaction term between feedback type and creative task was significant (p-value lower than 0.05), leading to the rejection of the null hypothesis for Hypothesis 2 (H0b). The main effect of feedback type was significant (p-value lower than

0.05), further rejecting the null hypothesis for Hypothesis 2. However, the effect of the creative task alone did not yield a significant result (p-value higher than 0.05), failing to reject the null hypothesis regarding the influence of task type on creativity. These results are meticulously documented to prepare for a thorough exploration in the Discussion section, where the implications of these findings will be considered in the broader context of existing literature on creativity and feedback mechanisms. This clear delineation of results provides a robust basis for subsequent interpretative analysis.

Regarding the third hypothesis (H0d), which examined the effect of any feedback versus no feedback, irrespective of the task, t-tests were conducted separately for story continuation and advertisement tasks. The results for the story continuation task showed a significant difference between specific feedback and no feedback, with a p-value lower than 0.05, rejecting the null hypothesis. The same was true for the comparison between general feedback and no feedback, with a p-value lower than 0.05. Conversely, for the advertisement task, the comparison between specific feedback and no feedback revealed a significant difference with a p-value lower than 0.05, rejecting the null hypothesis. However, the general feedback versus no feedback comparison did not yield a significant difference, with a p-value higher than 0.05, failing to reject the null hypothesis

7.3 Sentiment Analysis

In addition to examining semantic similarities to measure the originality of the participants' assignments, the study examines the depth of sentiment in creative outputs under all feedback conditions — specific, general, and no feedback (control group). This analysis was performed using two advanced RoBERTa models trained on distinct datasets: the cardiffnlp/twitter-roberta-base-sentiment-latest and siebert/sentiment-roberta-large-english, both sourced from HuggingFace and designed to analyze sentiment in English language text, thereby ensuring compatibility with the translated assignments of participants.

The sentiment analysis was stratified by task type—story continuation and advertisement copywriting for a computer mouse—across the three feedback modalities. For story continuation under the specific feedback condition, the twitter-based RoBERTa model revealed a sentiment distribution of 25.01% negative, 51.28% neutral, and 23.72% positive. In contrast, the general feedback condition showed a slightly more positive sentiment distribution with 19.05% negative, 53.4% neutral, and 27.55% positive. The control group exhibited the highest negative sentiment at 36.18% and the lowest positive sentiment at 13.2%. The sentiment analysis for the advertisement task under specific feedback displayed a predominantly positive sentiment (92.91%), significantly higher compared to the general feedback (75.5%) and control (85.22%) conditions.

Further depth in sentiment analysis using the siebert/sentiment-roberta-large-english model also highlighted significant variances. Specifically, the story continuation task under specific feedback exhibited a robustly positive sentiment at 85.72%, only slightly higher than the general feedback condition at 85.16%, suggesting minimal variation in sentiment positivity between these feedback types. However, the control condition for the same task showed a marked reduction in positive sentiment at 63.8%. For the advertisement task, both specific and general feedback conditions yielded nearly identical positive sentiment scores at 99.89%, overshadowing the control condition's also high but slightly lower score at 99.89%. Crucially, the t-test conducted to assess the fourth hypothesis—which posited that the average sentiment depth of creative outputs in the specific feedback condition is greater than that in the general feedback condition—comparing specific versus general feedback modalities yielded a p-value of 0.999. This indicates no statistically significant difference in sentiment depth between these conditions, thus failing to reject the null hypothesis (H0d).

Crucially, the t-test conducted across all sentiments comparing specific versus general feedback modalities yielded a p-value of 0.999, indicating no statistically significant difference in sentiment depth between these conditions. This outcome fails to reject the null hypothesis (H0d) of the fourth hypothesis, which posited that there is no significant difference in the average sentiment depth between creative outputs generated under the specific and general feedback conditions. Consequently, this result does not support the alternative hypothesis (H1d) that posited a greater average sentiment depth in the specific feedback condition, potentially indicating more creative elements.

7.4 Analysis of Semantic Quality

In order to evalaute the readability and complexity of creative outputs across different feedback modalities the Flesch-Kincaid Grade Level (FKG) and the Coleman-Liau Index (CLI) were used as measures. As mentioned in the theoretical framework these metrics provide insights into the textual quality influenced by the specificity and generality of the feedback provided during the creative tasks of story continuation and advertisement copywriting.

For the story continuation task, the average Flesch-Kincaid scores were 6.875 for specific feedback, 7.8 for general feedback, and 9.378 for the control group, indicating that texts produced with specific feedback were easiest to read, followed by general feedback, with the control group's outputs being the most complex.

In contrast, the Coleman-Liau Index scores presented a different trend: 13.046 for specific feedback, 12.017 for general feedback, and 11.008 for the control, suggesting that texts from the specific feedback group had the highest linguistic complexity, followed by those from the general and control groups. The advertisement task showed similar patterns in Flesch-Kincaid scores with 6.563 for specific feedback, 7.243 for general feedback, and 11.833 for the control, aligning with the story continuation task results in terms of readability. However, the CLI scores for the advertisement task were highest for the specific feedback at 14.054, followed by the control at 14.217, and general feedback at 12.903, indicating a higher complexity in the specific and control feedback texts compared to the general feedback. In testing the fifth hypothesis, which examined whether the readability and complexity of feedback influenced the semantic quality of the texts, t-test results provided mixed outcomes. The Flesch-Kincaid scores between specific and general feedback modalities did not reveal a statistically significant difference, with a p-value of 0.367, thus failing to reject the null hypothesis (H0e) that states no significant difference in readability levels between the specific and general feedback conditions. Conversely, the Coleman-Liau Index scores showed a significant difference, with a p-value of 0.008, suggesting that the texts from the specific feedback condition were indeed more complex, as indicated by higher CLI scores compared to those from the general feedback condition. This result supports the alternative hypothesis (H1e) that specific feedback is associated with a higher level of complexity, requiring a greater reading level for comprehension.

7.5 Summary of Findings

This study's findings give deep insights into how AI-generated feedback affects human creativity. The results supports the theory that precise, detailed feedback boosts creativity more efficiently than generic input, especially in tasks requiring high originality, such as advertising. THis also corresponds to the theories and concepts introduced in the theoretical framework of this study. However, the impact of input specificity on emotional depth appears to be minor, since sentiment analysis reveals no significant differences between feedback kinds. Furthermore, while particular input did not affect readability, it did increase textual complexity, implying an increase in the cognitive engagement necessary to understand such feedback. Collectively, these findings highlight the multiple impacts of feedback specificity on creativity, with important implications for the use of AI in creative processes.

8 Discussion

In the discussion of this thesis we aim to critically analyze the findings obtained from the experimental evaluation of feedback specificity provided by LLMs on human creativity. Following the detailed presentation of results, the discussion will interpret these findings, contextualize them within the existing literature, and explore their broader implications, thereby explaining the impact of feedback characteristics on creative outcomes.

8.1 Interpretation of findings

The study clearly showed that particular feedback considerably improves the creative process. The rejection of the null hypothesis for hypothesis 1 and 2 demonstrates that specific feedback performs better in enhancing creativity than broad feedback, especially in tasks that require a high level of originality. This improved semantic alignment shows that extensive feedback may help participants better grasp and satisfy creative benchmarks, maybe due to the richer and more relevant information it contains. In contrast, Hypothesis 3's findings suggest that any input, particular or generic, promotes creativity when compared to no feedback at all. This emphasizes the critical importance of feedback in encouraging creative production, demonstrating that simply getting feedback may improve creative engagement, regardless of its specifics.

The results of Hypothesis 4, which demonstrate no significant difference in sentiment depth between feedback types, suggest that feedback specificity promotes semantic and perceived creativity but has no meaningful effect on the emotional depth of creative outputs. This research suggests that additional elements, maybe inherent in the individual or task, may influence emotional engagement in creative pursuits.

Finally, Hypothesis 5 investigated the effects of feedback specificity on readability and text quality. The conflicting results—rejecting the null hypothesis for the Coleman-Liau Index but not for the Flesch-Kincaid Grade—show a complicated relationship between feedback specificity and text complexity. Specific feedback resulted in outputs with increased linguistic complexity without significantly reducing overall readability. Higher CLI scores indicate a sophisticated use of language that is nevertheless understandable, most likely owing to the participants' ability to digest and comprehend the context offered by particular feedback.

8.2 Link of Findings to Theory

The findings of this study establish an interesting connection to current cognitive theories claiming that comprehensive and contextually appropriate feedback improves learning and creativity. Specifically, the research confirms Hypotheses 1 and 2, which are consistent with educational theories emphasizing the value of accurate, actionable feedback in supporting cognitive growth and creative production. Furthermore, Hypothesis 3's findings reinforce the fundamental educational premise that feedback, regardless of specificity, serves as a stimulant for creativity, confirming its critical function in cognitive and creative processes. However, the findings from Hypotheses 4 and 5 provide a more nuanced picture, suggesting that, while particular feedback has a large impact on semantic and novelty components of creativity, its influence on emotional depth and readability is less strong. This intricacy poses a partial challenge to current ideas, showing that while feedback can improve cognitive aspects of creative work, its influence on emotional components may be more restricted, revealing an opportunity for theoretical growth.

8.3 Implications

Participants who received specific feedback not only spent more time on their submissions but also produced longer texts. This indicates a higher level of engagement and commitment, which likely contributed to the enhanced creativity observed. The increase in both time and word count in response to specific feedback suggests that when participants receive detailed, contextual guidance, they are more inclined to explore their creative capacities more fully. These results intuitively align with the hypothesis results, suggesting that specific feedback is not only more effective but also encourages a deeper investment in the creative process. Such profound impacts of specific feedback on engagement and creativity clearly underscore its practical significance. The implications of these findings are significant for educational technology developers and educators who utilize AI tools like LLMs to facilitate learning and creative expression. Specifically, the results advocate for the implementation of systems that can provide tailored, specific feedback to users, thereby potentially enhancing creative outcomes in academic and professional settings.Furthermore, identifying the bounds of feedback's influence on emotional depth might assist in designing better human-centered AI interfaces that cater to both the emotional and cognitive components of creativity.

8.4 Limitations

One major concern is the use of English translation for non-English source materials, which may have resulted in biases in both semantic and sentiment analysis. Furthermore, and this is why the translation was done, the sample size was rather small, and the distribution of participants was unequal between the two groups, with more participants in the control group. This unequal distribution, together with the variety of languages and analytical approaches used, has the potential to bias the results and restrict the findings' generalizability.

8.5 Further Research

Future studies should aim to replicate these findings with a larger, more evenly distributed sample size and within a monolingual context to eliminate translation biases and enhance the robustness of conclusions. Employing a consistent state-of-the-art SBERT model and exploring different AI models and feedback mechanisms could mitigate issues related to model discrepancies and clarify how various types of feedback influence creativity, motivation, and satisfaction. Additionally, a manual analysis of creative outputs using diverse metrics and assessment tools could provide deeper insights into the nuances of creativity that automated methods might overlook, offering a more nuanced understanding of how specific feedback influences different dimensions of creativity and text quality.

9 Conclusion

In conclusion, this thesis has provided a comprehensive examination of how the variation of feedback provided by LLMs influences the creativity and quality of human work. In response to the research question, "To what extent does the variation of feedback provided by the LLM influence the creativity and quality of human work?", the findings clearly demonstrate that feedback specificity drastically improves the creativity and quality of outputs. Specific feedback, as opposed to general or no feedback, results in greater semantic alignment with established creative norms and fosters more engaged and higher-quality creative performance.

These findings have significant implications for educational technology developers and educators, who advocate for the use of AI-driven systems that provide personalized, context-specific feedback to improve learning and creativity. Such feedback not only encourages large increases in creativity, but it also enhances the overall quality of the creative work. However, the study found that, while feedback specificity has a significant impact on semantic and creative quality, its impact on emotional depth is less strong, indicating that other factors may influence emotional elements of creativity.

Future research should try to build on these findings by using bigger, more varied samples and consistent methodology to better understand the subtle consequences of feedback specificity. This study not only adds to the academic conversation on AI, feedback, and creativity, but it also provides the framework for more effective and nuanced use of technology in nurturing human creativity. The fundamental impact of feedback variation, particularly in terms of specificity and detail, is therefore critical for improving creative outputs, corroborating the core premise of this study.

A Appendix

A.1 Creative survey tasks for specific feedback modality

The task description shown below is the creative task where the respondent is asked to write a creative story continuation. Once the participant is done the assignment should be copied and pasted to ChatGPT along with the pre-defined prompt in order to obtain detailed feedback.

> Schreibe eine kreative Fortsetzung zur folgenden Geschichte: Du bist Kapitän eines magischen Schiffs, dass durch die karibischen Meere umhersegelt als plötzlich ...
> Prompt für Feedback an ChatGPT: Sie sind ein erfahrener Kritiker, der darauf spezialisiert ist, detailliertes Feedback zur kreativen Arbeit von Menschen zu geben. Ihre Aufgabe ist es, dem Autor konstruktives und präzises Feedback zu seinem Text zu geben. Gehen Sie über oberflächliche Betrachtungen hinaus und konzentrieren Sie sich auf Überzeugungskraft, Sprachstil oder ähnliches. Betonen Sie Stärken und Schwächen im Text und bieten Sie konkrete Verbesserungsvorschläge sowie mögliche Richtungen für die Weiterentwicklung der Geschichte an. Ihr Ziel ist es, dem Autor ein tieferes Verständnis für die Qualität seines Werkes zu vermitteln und ihm zu helfen, sein kreatives Potenzial voll auszuschöpfen. Fertige niemals Text selbst an sondern gebe nur Feedback Hier ist der Text: ———
>
> ———————————————————————
>
> ————
>
> Kopieren Sie diesen Prompt mit Ihrem Text und fügen Sie ihn bei ChatGPT ein für Feedback. Schauen Sie sich das Feedback von ChatGPT genau an und überlegen sie wie sie Ihren Text überarbeiten können. Überarbeiten Sie Ihren Text wenigstens einmal
> Gerne können Sie, das Eingabefeld dazu nutzen um einen ersten Text zu schreiben, den sie dann bei ChatGPT einfügen.
> Wir sind nur an ihrer finalen Lösung interessiert, und nicht was Sie vor dem Feedback geschrieben haben. Schreiben Sie bitte hier Ihre überarbeitete und finale Endlösung:

The task description shown below is the creative task where the respondent is asked to write creative and compelling advertisement copy on a computer mouse. Once the participant is done the assignment should be copied and pasted to Chat-

GPT along with the pre-defined prompt in order to obtain detailed feedback.

> Schreiben Sie einen überzeugenden Marketingtext für eine Computermaus. Ihr Text soll die einzigartigen Eigenschaften und Vorteile der Computermaus hervorheben. Verwenden Sie kreative Sprache, überzeugende Argumente (z.B. ergonomisches Design, Präzision, innovative Technologie) und detaillierte Beschreibungen, um die Aufmerksamkeit der Kunden zu gewinnen und sie zum Kauf zu motivieren. Seien Sie kreativ und beschreiben Sie, dass es das perfekte Werkzeug für produktives Arbeiten ist.
> Prompt für Feedback an ChatGPT: Sie sind ein erfahrener Kritiker, der darauf spezialisiert ist, detailliertes Feedback zur kreativen Arbeit von Menschen zu geben. Ihre Aufgabe ist es, dem Autor konstruktives und präzises Feedback zu seinem Text zu geben. Gehen Sie über oberflächliche Betrachtungen hinaus und konzentrieren Sie sich auf spezifische Aspekte wie Charakterentwicklung, Handlungsführung, Sprachstil und thematische Kohärenz. Identifizieren Sie Stärken und Schwächen im Text und bieten Sie konkrete Verbesserungsvorschläge sowie mögliche Richtungen für die Weiterentwicklung der Geschichte an. Ihr Ziel ist es, dem Autor ein tieferes Verständnis für die Qualität seines Werkes zu vermitteln und ihm zu helfen, sein kreatives Potenzial voll auszuschöpfen. Fertige niemals Text selbst an sondern gebe nur Feedback. Hier ist der Text: ———————————————————— Kopieren Sie diesen Prompt mit Ihrem Text und fügen Sie ihn bei ChatGPT ein für Feedback. Schauen Sie sich das Feedback von ChatGPT genau an und überlegen sie wie sie Ihren Text überarbeiten können. Überarbeiten Sie Ihren Text wenigstens einmal
> Gerne können Sie, das Eingabefeld dazu nutzen um einen ersten Text zu schreiben, den sie dann bei ChatGPT einfügen.
> Wir sind nur an ihrer finalen Lösung interessiert, und nicht was Sie vor dem Feedback geschrieben haben. Schreiben Sie bitte hier Ihre überarbeitete und finale Endlösung:

A.2 Creative survey tasks for general feedback modality

Similarly to the specific feedback modality above the general feedback modality tasks are just slightly different. The difference lies in the prompts that were designed to only provide the participant with broad and general feedback. The task is except for the prompt identical.

Below is the task for the story continuation task.

Schreibe eine kreative Fortsetzung zur folgenden Geschichte: Du bist Kapitän eines magischen Schiffs, dass durch die karibischen Meere umhersegelt als plötzlich ... Prompt für Feedback an ChatGPT: Sie sind ein Bewerter, der dafür zuständig ist, allgemeines Feedback zur kreativen Arbeit von Menschen zu geben. Ihre Aufgabe ist es, dem Autor ein allgemeines und eher vages Feedback zu seinem Text zu geben. Seien Sie nicht zu spezifisch, sondern geben Sie eher allgemeines Feedback und Dinge, die der Autor in Betracht ziehen könnte oder die er verbessern kann. Sie können nicht beurteilen, ob der kreative Text fertig ist. Nutzen Sie Ihr Wissen über kreatives Schreiben, um dem Autor allgemeine und vage Hinweise zu geben, wohin er gehen könnte. Schreiben Sie niemals einen Text für den Autor, sondern geben Sie nur Feedback. Hier ist der Text: ______________________________ Kopieren Sie diesen Prompt mit Ihrem Text und fügen Sie ihn bei ChatGPT ein für Feedback. Schauen Sie sich das Feedback von ChatGPT genau an und überlegen sie wie sie Ihren Text überarbeiten können. Überarbeiten Sie Ihren Text wenigstens einmal. Gerne können Sie, das Eingabefeld dazu nutzen um einen ersten Text zu schreiben, den sie dann bei ChatGPT einfügen. Wir sind nur an ihrer finalen Lösung interessiert, und nicht was Sie vor dem Feedback geschrieben haben. Schreiben Sie bitte hier Ihre überarbeitete und finale Endlösung:

And below here is the task for the advertisement copy for the general feedback modality group.

Schreiben Sie einen überzeugenden Marketingtext für eine Computermaus. Ihr Text soll die einzigartigen Eigenschaften und Vorteile der Computermaus hervorheben. Verwenden Sie kreative Sprache, überzeugende Argumente (z.B. ergonomisches Design, Präzision, innovative Technologie) und detaillierte Beschreibungen, um die Aufmerksamkeit der Kunden zu gewinnen und sie zum Kauf zu motivieren. Seien Sie kreativ und beschreiben Sie, dass es das perfekte Werkzeug für produktives Arbeiten ist. Prompt für Feedback an ChatGPT: Sie sind ein Bewerter, der dafür zuständig ist, allgemeines Feedback zur kreativen Arbeit von Menschen zu geben. Ihre Aufgabe ist es, dem Autor ein allgemeines und eher vages Feedback zu seinem Text zu geben. Seien Sie nicht zu spezifisch, sondern geben Sie eher allgemeines Feedback und Dinge, die der Autor in Betracht ziehen könnte oder die er verbessern kann. Sie können nicht beurteilen, ob der kreative Text fertig ist. Nutzen Sie Ihr Wissen über kreatives Schreiben, um dem Autor allgemeine und vage Hinweise zu geben, wohin er gehen könnte. Schreiben Sie niemals einen Text für den Autor, sondern geben Sie nur Feedback. Hier ist der Text: ——————————————— Kopieren Sie diesen Prompt mit Ihrem Text und fügen Sie ihn bei ChatGPT ein für Feedback. Schauen Sie sich das Feedback von ChatGPT genau an und überlegen sie wie sie Ihren Text überarbeiten können. Überarbeiten Sie Ihren Text wenigstens einmal. Gerne können Sie, das Eingabefeld dazu nutzen um einen ersten Text zu schreiben, den sie dann bei ChatGPT einfügen. Wir sind nur an ihrer finalen Lösung interessiert, und nicht was Sie vor dem Feedback geschrieben haben. Schreiben Sie bitte hier Ihre überarbeitete und finale Endlösung:

A.3 Control group

The control group served as a baseline for comparison and therefore respondents did not receive any feedback from the LLM ChatGPT. Therefore those respondents only saw the task without a pre-defined prompt and without instructions to use ChatGPT.

Below is the story continuation task for the control group.

> Schreibe eine kreative Fortsetzung zur folgenden Geschichte: Du bist Kapitän eines magischen Schiffs, dass durch die karibischen Meere umhersegelt als plötzlich ...

And below here is the task for the advertisement copy for the control group.

> Schreiben Sie einen überzeugenden Marketingtext für eine Computermaus. Ihr Text soll die einzigartigen Eigenschaften und Vorteile der Computermaus hervorheben. Verwenden Sie kreative Sprache, überzeugende Argumente (z.B. ergonomisches Design, Präzision, innovative Technologie) und detaillierte Beschreibungen, um die Aufmerksamkeit der Kunden zu gewinnen und sie zum Kauf zu motivieren. Seien Sie kreativ und beschreiben Sie, dass es das perfekte Werkzeug für produktives Arbeiten ist.

A.4 GitHub Repository

The Python code that I programmed in order to analyze the data and conduct the hypotheses tests are visible under the following link to my public GitHub repository for this bachelor thesis: https://github.com/stevenmkhitarian/bachelor_thesis.git

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Steven Mkhitarian

Köln, den 17.04.2024

References

Abdullah, M., Madain, A., & Jararweh, Y. (2022). Chatgpt: Fundamentals, applications and social impacts. *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 1-8. doi: 10.1109/SNAMS58071.2022.10062688

Albert, R. S., & Runco, M. (1998). A history of research on creativity. , 16-32. doi: 10.1017/CBO9780511807916.004

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, *45*, 357-376. doi: 10.1037/0022-3514.45.2.357

Amabile, T. M. (1996). *Creativity and innovation in organizations* (Vol. 5). Harvard Business School Boston.

Amabile, T. M. (2012). Componential theory of creativity.

doi: 10.4135/9781452276090.n42

Amabile, T. M., & Pillemer, J. (2012). Perspectives on the social psychology of creativity. *Journal of Creative Behavior*, *46*, 3-15. doi: 10.1002/JOCB.001

Anantrasirichai, N., & Bull, D. (2020). Artificial intelligence in the creative industries: a review. *Artificial Intelligence Review*, *55*, 589-656. doi: 10.1007/s10462-021-10039-7

Anderson, N., Potočnik, K., & Zhou, J. (2014). Innovation and creativity in organizations. *Journal of Management*, *40*, 1297 - 1333. doi: 10.1177/0149206314527128

Atkinson, D. P., & Barker, D. R. (2023). Ai and the social construction of creativity. *Convergence*, *29*, 1054 - 1069. doi: 10.1177/13548565231187730

Barbieri, F., Camacho-Collados, J., Neves, L., & Espinosa-Anke, L. T. (2020). Unified benchmark and comparative evaluation for tweet classification. arxiv 2020. *arXiv preprint arXiv:2010.12421*.

Barile, S., Bassano, C., Piciocchi, P., Vito, P., & Spohrer, J. C. (2022). Algorithms and human creativity: Threats or opportunities?a literature review. *The Human Side of Service Engineering*. doi: 10.54941/ahfe1002563

Boden, M. (1998). Creativity and artificial intelligence. *Artif. Intell.*, *103*, 347-356. doi: 10.1016/S0004-3702(98)00055-1

Camacho-collados, J., Rezaee, K., Riahi, T., Ushio, A., Loureiro, D., Antypas, D., ... others (2022, December). TweetNLP: Cutting-edge natural language processing for social media. In *Proceedings of the 2022 conference on empirical methods in natural language processing: System demonstrations* (pp. 38–49). Abu Dhabi, UAE: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/2022.emnlp-demos.5`

Cetinic, E., & She, J. (2021). Understanding and creating art with ai: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *18*, 1 - 22. doi: 10.1145/3475799

Chen, Z., & Chan, J. (2023). Large language model in creative work: The role of collaboration modality and user expertise. *Available at SSRN 4575598*.

Choudhery, S., Xi, Y., Chen, H., Aboul-Fettouh, N., GoldenMerry, Y. L., Ho, C., ... Goudreau, S. (2020). Readability and quality of online patient education material on websites of breast imaging centers. *Journal of the American College of Radiology : JACR*. doi: 10.1016/j.jacr.2020.04.016

Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal*, *18*, 391 - 404. doi: 10.1207/s15326934crj1803_13

Dacey, J. (2011). Historical conceptions of creativity. , 608-616. doi: 10.1016/B978-0-12-375038-9.00042-X

Das, P., & Varshney, L. (2022). Explaining artificial intelligence generation and creativity: Human interpretability for novel ideas and artifacts. *IEEE Signal Processing Magazine*, *39*, 85-95. doi: 10.1109/MSP.2022.3141365

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, *125*(6), 627.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... Williams, M. D. (2019). Artificial intelligence (ai): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*. doi: 10.1016/J.IJINFOMGT.2019.08.002

Farooq, U., & Grudin, J. (2016). Human-computer integration. *Interactions*, *23*, 26 - 32. doi: 10.1145/3001896

Fischer, G. (2001). User modeling in human–computer interaction. *User Modeling and User-Adapted Interaction*, *11*, 65-86. doi: 10.1023/A:1011145532042

Franceschelli, G., & Musolesi, M. (2023). On the creativity of large language models. *arXiv preprint arXiv:2304.00008*.

Furman, J., & Seamans, R. C. (2018). Ai and the economy. *Innovation Policy and the Economy*, *19*, 161 - 191. doi: 10.1086/699936

Gobet, F., & Sala, G. (2019). How artificial intelligence can help us understand human creativity. *Frontiers in Psychology*, *10*. doi: 10.3389/fpsyg.2019.01401

Gutbezahl, J., & Averill, J. (1996). Individual differences in emotional creativity as manifested in words and pictures. *Creativity Research Journal*, *9*, 327-337. doi: 10.1207/S15326934CRJ0904_4

Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2023). More than a feeling: Accuracy and application of sentiment analysis. *International Journal of Research in Marketing*, *40*(1), 75-87. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167811622000477 doi: https://doi.org/10.1016/j.ijresmar.2022.05.005

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 112 - 81. doi: 10.3102/003465430298487

He, V. F., Shrestha, Y., Puranam, P., & Miron-Spektor, E. (2023). Searching together: A theory of human-ai co-creativity. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4603650

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, *2*, 230 - 243. doi: 10.1136/svn-2017-000101

Jordan, M. I., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*, 255 - 260. doi: 10.1126/science.aaa8415

Li, W., Li, X., Huang, L., zhen Kong, X., Yang, W., Wei, D., ... Liu, J. (2015). Brain structure links trait creativity to openness to experience. *Social cognitive and affective neuroscience*, *10 2*, 191-8. doi: 10.1093/scan/nsu041

Liu, M.-Y., Huang, X., Yu, J., Wang, T.-C., & Mallya, A. (2020). Generative adversarial networks for image and video synthesis: Algorithms and applications. *Proceedings of the IEEE*, *109*, 839-862. doi: 10.1109/JPROC.2021.3049196

Loureiro, D., Barbieri, F., Neves, L., Espinosa Anke, L., & Camacho-collados, J. (2022, May). TimeLMs: Diachronic language models from Twitter. In *Proceedings of the 60th annual meeting of the association for computational linguistics: System demonstrations* (pp. 251–260). Dublin, Ireland: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/2022.acl-demo.25` doi: 10.18653/v1/2022.acl-demo.25

Lubart, T., & Getz, I. (1997). Emotion, metaphor, and the creative process. *Creativity Research Journal*, *10*, 285-301. doi: 10.1207/S15326934CRJ1004_1

Maravilhas, S. (2015). Creativity, invention, and innovation. , 4071-4079. doi: 10.4018/978-1-4666-5888-2.CH401

Marchesi, M. (2017). Megapixel size image creation using generative adversarial networks. *ArXiv*, *abs/1706.00082*.

Morriss-Kay, G. M. (2010). The evolution of human artistic creativity. *Journal of anatomy*, *216*(2), 158–176.

Norton, D., Heath, D., & Ventura, D. (2013). Finding creativity in an artificial artist. *Journal of Creative Behavior*, *47*, 106-124. doi: 10.1002/JOCB.27

Pope, R. (2005). Creativity: Theory, history, practice.

doi: 10.4324/9780203695319

Reimers, N., & Gurevych, I. (2019, 11). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing.* Association for Computational Linguistics. Retrieved from `https://arxiv.org/abs/1908.10084`

Rezwana, J., & Maher, M. (2022). Designing creative ai partners with cofi: A framework for modeling interaction in human-ai co-creative systems. *ACM Transactions on Computer-Human Interaction*, *30*, 1 - 28. doi: 10.1145/3519026

Ryan, R., & Deci, E. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25 1*, 54-67. doi: 10.1006/CEPS.1999.1020

Sarooghi, H., Libaers, D., & Burkemper, A. C. (2015). Examining the relationship between creativity and innovation: A meta-analysis of organizational, cultural, and environmental factors . *Journal of Business Venturing*, *30*, 714-731. doi: 10.1016/J.JBUSVENT.2014.12.003

Sternberg, R., & Lubart, T. (1998). The concept of creativity: Prospects and paradigms. , 3-15. doi: 10.1017/CBO9780511807916.003

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, *185*(4157), 1124–1131.

Webster, S. (2019). Readability index operation in research. , *9*, 25-26. doi: 10.37591/JONSP.V9I1.1509

Yuan, A., Coenen, A., Reif, E., & Ippolito, D. (2022). Wordcraft: Story writing with large language models. *27th International Conference on Intelligent User Interfaces*. doi: 10.1145/3490099.3511105

Yusof, R., Amin, R., Zainudin, R., & Baker, O. F. (2004). Human computer interaction (hci) - aspect in developing information access system. *2004 IEEE Region 10 Conference TENCON 2004.*, *B*, 387-390 Vol. 2. doi: 10.1109/TENCON.2004.1414613

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., … rong Wen, J. (2023). A survey of large language models. *ArXiv*, *abs/2303.18223*. doi: 10.48550/arXiv.2303.18223

Zhou, J. (1998). Feedback valence, feedback style, task autonomy, and achievement orientation: Interactive effects on creative performance. *Journal of Applied Psychology*, *83*, 261-276. doi: 10.1037/0021-9010.83.2.261