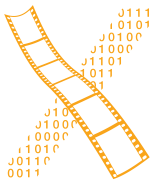


Chair for Computer Science 10 (Media Computing and Human-Computer Interaction)

RWTH AACHEN UNIVERSITY

Exploring the Impact of Generative AI Tools on Creative Co-located Group Processes

Master's Thesis at the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Raghad Zaghal

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Ulrik Schroeder

Registration date: 17.01.2025
Submission date: 24.03.2025

Eidesstattliche Versicherung
Declaration of Academic Integrity

Name, Vorname/Last Name, First Name

Matrikelnummer (freiwillige Angabe)
Student ID Number (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/ Masterarbeit* mit dem Titel
I hereby declare under penalty of perjury that I have completed the present paper/bachelor's thesis/master's thesis* entitled

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt; dies umfasst insbesondere auch Software und Dienste zur Sprach-, Text- und Medienproduktion. Ich erkläre, dass für den Fall, dass die Arbeit in unterschiedlichen Formen eingereicht wird (z.B. elektronisch, gedruckt, geplottet, auf einem Datenträger) alle eingereichten Versionen vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.
independently and without unauthorized assistance from third parties (in particular academic ghostwriting**.** I have not used any other sources or aids than those indicated; this includes in particular software and services for language, text, and media production. In the event that the work is submitted in different formats (e.g. electronically, printed, plotted, on a data carrier), I declare that all the submitted versions are fully identical. I have not previously submitted this work, either in the same or a similar form to an examination body.

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen/Please delete as appropriate

Belehrung:
Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt
Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.
§ 156 StGB (German Criminal Code): False Unsworn Declarations
Whosoever before a public authority competent to administer unsworn declarations (including Declarations of Academic Integrity) falsely submits such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment for a term not exceeding three years or to a fine.
§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt
(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.
(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.
§ 161 StGB (German Criminal Code): False Unsworn Declarations Due to Negligence
(1) If an individual commits one of the offenses listed in §§ 154 to 156 due to negligence, they are liable to imprisonment for a term not exceeding one year or to a fine.
(2) The offender shall be exempt from liability if they correct their false testimony in time. The provisions of § 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:
I have read and understood the above official notification:

Ort, Datum/City, Date

Unterschrift/Signature

Contents

List of Figures and Tables

Abstract

Generative artificial intelligence (AI) is increasingly becoming a fundamental tool in collaborative and creative work, affecting how teams and individuals develop and refine ideas. Although previous studies have examined the influence of AI on individual creativity, its influence on co-located group collaboration remains underexplored. This thesis examines how generative AI tools affect co-located group creativity, collaboration patterns, and creative outputs in two studies. We conducted our first study, using within-subjects design, with groups of three, who complete marketing-oriented creative tasks in two conditions: (1) with AI assistance and (2) without AI. The second study follows, in which external observers evaluate the creative outputs of the first experiment. The findings of this thesis are relevant for human-AI interaction, creative industries, and co-located group work in educational and professional settings. The research questions addressed during the thesis explore how AI influences co-located collaboration among groups, how AI influences creative outputs, and participants' opinions on collaborating while using generative AI tools.

Überblick

Generative künstliche Intelligenz (KI) wird immer mehr zu einem grundlegenden Werkzeug für die Zusammenarbeit und die kreative Arbeit, da sie die Art und Weise beeinflusst, wie Teams und Einzelpersonen Ideen entwickeln und umsetzen. Obwohl frühere Studien den Einfluss von KI auf die individuelle Kreativität untersucht haben, ist ihr Einfluss auf die Zusammenarbeit in Gruppen noch nicht ausreichend erforscht. In dieser Arbeit wird in zwei Studien untersucht, wie sich generative KI-Tools auf die Gruppenkreativität, die Kooperationsmuster und die kreativen Ergebnisse auswirken. Um dies zu erforschen, haben wir in der ersten Studie ein Within-Subjects-Design mit Dreiergruppen durchgeführt, die marketingorientierte kreative Aufgaben unter zwei Bedingungen bearbeiten: (1) mit KI-Unterstützung und (2) ohne KI. Die zweite Studie folgt, in der externe Beobachter die kreativen Ergebnisse des ersten Experiments bewerten. Die Ergebnisse dieser Arbeit sind relevant für die Interaktion zwischen Mensch und KI, die Kreativwirtschaft und die gemeinsame Gruppenarbeit in Ausbildung und Beruf. Die Forschungsfragen, die in dieser Arbeit behandelt werden, befassen sich damit, wie KI die Zusammenarbeit zwischen Gruppen an einem Ort beeinflusst, wie KI die kreativen Ergebnisse beeinflusst und wie die Teilnehmer über die Zusammenarbeit unter Verwendung generativer KI-Tools denken.

Acknowledgments

I would like to express my sincere gratitude to Prof. Dr. Jan Borchers and Prof. Dr. Ulrik Schroeder for reviewing my thesis and providing valuable feedback. A special thank you to my advisor, Paul Preuschoff, for giving me the opportunity to write my thesis at Chair i10, and for his continuous support, guidance, and encouragement throughout the past months. I am also grateful to Chair i10 for welcoming me as a thesis student and providing me with an inspiring research environment.

A heartfelt thank you to all the participants who volunteered for my user study and survey. Your time and contributions were invaluable to this research.

Finally, I extend my deepest appreciation to my fiancé, Nasim, and my parents, Iman Jaber and Dr. Raid Zaghal, whose unwavering support, love, and belief in the power of education have enabled me to pursue my studies abroad and achieve my academic goals. I am forever grateful for their encouragement and the opportunities they have given me.

Conventions

Throughout this thesis we use the following conventions:

- The thesis is written in American English.
- The first person is written in plural form.
- Unidentified third persons are described in female form.

Short excursuses are set off in colored boxes.

EXCURSUS:
Excursuses are set off in orange boxes.

Where appropriate, paragraphs are summarized by one or two sentences that are positioned at the margin of the page.

This is a summary of a paragraph.

Chapter 1

Introduction

"Generative models are a key enabler of machine creativity, allowing machines to go beyond what they've seen before and create something new."

—Ian Goodfellow, Machine Learning Researcher and Creator of GANs

Artificial intelligence definition

Artificial Intelligence (AI) is a concept that has been defined in numerous ways. It is often defined as the ability of machines to think and act like human beings, to reason logically, or to function as rational agents that can make decisions to achieve specific goals [Peter and Intelligence, 2021]. However, in recent years, generative AI has reshaped many different fields, ranging from computer science to creative arts [Bandi et al., 2023]. Generative AI is a type of artificial intelligence that is capable of producing seemingly original content, such as text, images, or audio, based on patterns and data from their training sets [Feuerriegel et al., 2024].

Generative AI is transforming creative and collaborative work

The increased popularity of generative AI tools, such as DALL-E and GPT-4, is transforming the way people create media, communicate and collaborate with each other [Feuerriegel et al., 2024]. These tools are now becoming central in collaborative and creative settings, which makes it increasingly important to study their impact on group dynamics and group creativity [Tomić et al., 2023; Zhang et al., 2024].

Generative AI can enhance creative work but it introduces unique challenges

Studying the effects of generative AI on creativity is vital because creativity is the essence of human innovation [Nijstad, 2015]. Creativity drives how we approach problem-solving, write stories, and create art. As societies and industries evolve, creativity remains an essential skill, especially in contexts where new ideas and collaborative efforts are valued. Generative AI tools push the limits of creativity by providing new ways for humans to explore their creative abilities, however, generative AI tools have also introduced new and unique challenges. They are reshaping what it means to innovate, by redefining the dynamics of human creativity and human-computer interaction, allowing new forms of collaboration that blend human intuition with AI-driven efficiency[Luan, 2024].

Research on AI's role in group creative processes

Although significant research has been done on the influence of AI on individual creativity, such as Luan [2024]'s "The New Creative Alliance: Investigating the Dynamics of Human-AI Collaboration in Creative Endeavours", less attention has been paid to its role in co-located group settings. Research shows that AI tools can reduce the cognitive workload of its users by automating repetitive or data-intensive tasks, which enables teams to focus more on the strategic and creative aspects of their work [Braun et al., 2023]. Other studies highlight the importance of user autonomy and attitudes, in which they found that user's positive attitudes toward AI and the freedom to use it flexibly can enhance the quality of their collaboration with AI [Bezrukova et al., 2023]. However, the majority of existing research has focused on virtual collaboration or individual work, leaving a notable gap in understanding how AI influences creative processes in co-located group settings.

1.1 Outline

This thesis explores how AI impacts group creativity in co-located settings

The aim of this thesis is to address the aforementioned research gap by exploring the impact of generative AI tools on creative processes of groups collaborating in the same physical location. The research is particularly relevant in professional and educational environments where teams often collaborate on creative projects. As generative AI

tools become increasingly accessible and embedded in everyday workflows, understanding their impact on group dynamics and creativity is vital for developing effective and ethical practices and tools [Jansen and Sklar, 2021; Yusa et al., 2022].

This thesis explores the following research questions (RQs):

RQ1 How do generative AI tools impact the creative process and outcomes in co-located group settings?

- **RQ1.1** How do AI tools influence collaboration dynamics in the creative process of the group?
- **RQ1.2** How do participants perceive the role of AI in creative group work?
- **RQ1.3** How does AI impact how group members perceive their group's creative result?
- **RQ1.4** How does AI impact how others perceive a group's creative result?

Concept of Study 1

To explore these questions, we performed two complementary studies. Study 1 involves a lab study with groups of three participants, who are asked to work on creative tasks in marketing. This study employs a within-subject design, in which the first condition involves using a generative AI tools (ChatGPT), while the other condition requires participants to complete their tasks without any AI assistance. During the study, data is collected through questionnaires, video recordings, and observational notes to analyze how AI tools influence the creative process, group dynamics, and task outcomes.

The thesis includes a lab study on AI-assisted group creativity and a survey (Study 2) evaluating the outputs from Study 1

Study 2 involves a survey in which external participants evaluate and compare the creative tasks generated by the participants during the lab study. This survey aims to assess how AI-assisted and non-AI-assisted outputs are perceived by external raters, providing our research with additional insights into the impact of generative AI tools on creative outputs.

The findings of this thesis aim to contribute to the broader understanding of how participants interact with generative

Thesis contributes insights on generative AI's impact on group dynamics and creative tasks

AI tools in collaborative settings, specifically for performing creative tasks. By identifying how these tools influence group dynamics and influence outcomes, the research provides valuable insights for designing future AI tools and determining how, when, and to what extent AI should be integrated into group collaboration, or whether it should be incorporated at all. Furthermore, outcomes from this research can contribute to improving collaborative processes in educational and professional contexts and driving innovation in creative industries.

Thesis structure summarized

The remaining structure of the thesis continues with Chapter 2, in which we review relevant literature on generative AI, creativity, and collaborative dynamics, and in Chapter 3, we highlight our research questions. In the following Chapter 4, we thoroughly discuss the process and methodology of the first user study, including the details on the study design, pilot study, finalized structure, data collection and results. The same is repeated for Study 2 in Chapter 5. The subsequent Chapter 6 presents the discussion and evaluation for both studies, along with a discussion of limitations. Finally, Chapter 7 concludes the thesis with a summary and directions for future research.

Chapter 2

Related Work

2.1 Generative AI: Foundations and Capabilities

Generative AI is advancing rapidly

Generative AI has transformed many fields such as computer vision, natural language processing, and creative arts. Its applications include image and text generation, music composition, and chat-bots [Bandi et al., 2023]. According to Bandi et al. [2023], this fast growth has been made possible with the ongoing development of deep learning techniques and access to extensive datasets. In their article *the Power of Generative AI*, Bandi et al. [2023] review the requirements, models, input and output formats of generative AI, as well as its evaluation metrics and challenges.

Different AI models specialize in various generative tasks

Generative AI models are classified into many types. Firstly, there are Variational Autoencoders (VAEs), which are often used for tasks requiring compressed data representations. There are also Generative Adversarial Networks (GANs), consisting of a generator and a discriminator that learn together to create realistic outputs [Goodfellow et al., 2014; Wang et al., 2017]. Transformers are models that capture global dependencies, often used for text or sequence generation. Meanwhile, diffusion models create

high-quality images, and hybrid models combine multiple approaches for diverse generation capabilities.

Evaluating AI involves quality, ethics, and computational cost [Bandi et al., 2023]

Bandi et al. [2023] also review the different types of input and output formats in generative AI. Common inputs include images, text, audio, and tabular data, while outputs vary from images and text to audio and structured data. The formats depend on applications, such as images for GANs or text for language models. In addition, Bandi et al. [2023] discuss how generative AI can be evaluated based on output quality, performance, and usability, before addressing the ethical challenges of authorship, originality, and data privacy. They also note that model training requires extensive resources, and that generating high-quality outputs can be computationally expensive. They also mention that evaluating AI is often subjective, especially for creative outputs [Bandi et al., 2023].

AI-driven art challenges traditional notions of authorship

Expanding the scope of generative AI, Yusa et al. [2022] review six case studies exploring how artificial intelligence can be integrated into artistic works. For instance, *The Next Rembrandt* replicates an artist's style by analyzing previous works, which raises questions about authorship. Another example, *Pulse Room*, uses sensors to convert viewers' heartbeats into light patterns, consequently emphasizing the connection between human presence and technology. Lastly, *Memoirs from Latent Space Study II* by Refik Anadol uses deep neural networks to explore the intersection of memories and technology [Yusa et al., 2022]. An example can be seen in Figure 2.1, which depicts a neural network-generated visualization from Refik Anadol's *Memoirs from Latent Space Study II (2019)*. This image shows an abstract interpretation of memory and experience.

2.2 AI in Creative Processes

In another paper on AI and art, "Too Late to be Creative? AI-Empowered Tools in Creative Processes" by Hwang [2022], the author categorizes AI tools into four groups: Editors, Transformers, Blenders, and Generators, and discusses co-creativity between humans and AI. Ac-

Figure 2.1: "Memoirs from Latent Space Study II" (2019) Yusa et al. [2022]

cording to Hwang [2022], the creative process consists of four stages: Q&A (defining the problem and gathering information), wandering (brainstorming), hands-on (generating and combining ideas), and camera-ready (evaluating and presenting final ideas). AI tools mostly support the hands-on and camera-ready stages but offer little help for Q&A or wandering. Hwang [2022]'s paper also examines human-AI interaction (HAII) challenges. *Divergent* challenges come from unpredictable outputs, while *convergent* challenges involve skepticism and mistrust over AI decision-making, and *collaborative* challenges relate to role establishment and users' fear of replacement by AI. Hwang [2022] emphasizes that future AI tools should also support early-stage ideation and promote transparency, creative autonomy, and collaborative features.

AI tools support later creative stages but struggle with early ideation and brainstorming

AI improves visual media design with advanced algorithms, but further research is needed for full integration

Additionally, Liu et al. [2021] explore how AI adds innovation to traditional color and image usage in visual media design. They define *AI Visual Media Communication Design*, which uses graphics, text, and machine learning for enhanced visuals, improved precision, and user perception. Liu et al. [2021] also review algorithms for design optimization, such as *Kalman Filtering* for noise reduction, and the *Gray Expected Image Algorithm* for balancing gray levels. They also mention how packaging and film/media industries have adopted AI quite actively, and report high satisfaction levels. Liu et al. [2021] conclude that, although still in early stages, integrating AI into visual communication design demands ongoing research to refine these technologies and fully realize their potential.

AI streamlines graphic design but raises concerns about creative autonomy, ethics, and skill redundancy

In another paper about AI and graphic design, Tomić et al. [2023]'s "Artificial Intelligence in Graphic Design" discusses how AI integration supports generative design, generating logos, as well as image and color manipulation. They explain that AI can be used for automating repetitive tasks like resizing and color matching to enable quicker and more optimal iterations. However, Tomić et al. [2023] highlight creative autonomy as a large concern. They raise questions about whether AI could replace designers instead of serving as a complementary tool. They also stress ethical implications, like authorship and copyright, since AI often relies on pre-existing data that was created by humans. Another issue they mention is skill redundancy, meaning that traditional design skills risk becoming obsolete. Tomić et al. [2023] conclude that AI should enhance rather than replace human creativity, and that ethical guidelines are crucial, along with a balanced integration of AI-generated work.

2.3 Human-AI Collaboration in Creative Workflows

Additionally, in another publication titled "Exploring Co-creative Drawing Workflows", Jansen and Sklar [2021] explore the topic of AI supporting human artists in collab-

orative drawing workflows. Their study involved professional illustrators and art students, and they aimed to investigate artists' attitudes toward AI as a collaborative tool. A key finding was that artists preferred AI for suggestions when they had artistic blocks. The study emphasizes that co-creative systems should not overshadow artists' input, but rather enhance it. Artists expressed interest in using AI for repetitive tasks like texturing and layout, and they preferred AI as a collaborative partner rather than a tutor. Another notable finding is the artists' preference for physical art mediums, due to tactile feedback. However, some were skeptical about AI, seeing it as a threat to their creative autonomy and lacking depth [Jansen and Sklar, 2021]. Their study concludes that AI can be a useful medium for creating art, provided it is designed with artists' autonomy in mind. They propose co-creative drawing systems that enhance productivity without overshadowing human input.

Artists value AI for suggestions and repetitive tasks but fear loss of creative autonomy

A significant inspiration for this thesis was drawn from *The New Creative Alliance: Investigating the Dynamics of Human-AI Collaboration in Creative Endeavors*, particularly the **third chapter** titled *Learning from AI or Human? A Comparative Study of Creative Outputs in Human-AI Collaboration versus Human-Only Conditions Across 10 Rounds of Creativity Tasks* by Luan [2024]. In this chapter of her dissertation, Luan [2024] compares human-AI versus human-only teams working through 10 creative tasks. One group solved tasks without AI, while the other used generative AI. The objectives included examining the influence of long-term human-AI collaboration on creativity, comparing creativity levels, and identifying interaction patterns. The study also included individual creativity tests, grouping subjects by high or low creativity, and assessing personality traits via the Big Five. Participants then proposed solutions to given problems, and Luan [2024] used the consensus assessment technique (CAT), with three independent evaluators scoring the creativity of each task's solution. We drew inspiration from this dissertation for our first user study, in which groups worked on creative tasks in an AI and No AI condition, however, some key differences between these two studies is that our study is a collaborative group study, which employs within-subject design. We also did not mea-

This thesis draws inspiration from a study comparing AI-assisted and human-only creative tasks

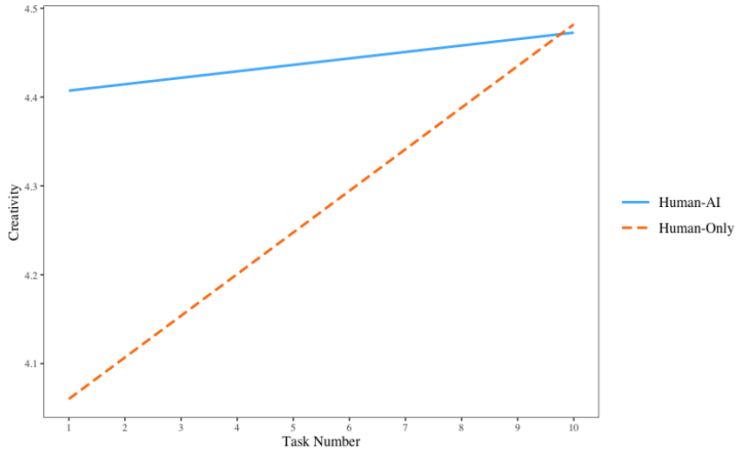


Figure 2.2: The interaction plot showing the effect of task number and condition on creativity. Luan [2024]

sure the creativity levels of the outputs by professional creativity evaluators, but with a normal survey study.

The findings of Luan [2024]'s study show that participants working with AI experienced an initial boost in creativity compared to those working alone. However, users working alone improved over time. This trend was not observed in the human-AI group. The difference can be seen in Figure 2.2. More observations revealed social loafing in human-AI settings. Meaning that users who perceived AI as a partner often relied too much on its input, and as a result exploring their personal creativity less [George, 1992]. Luan [2024] also found that more creative individuals benefited more from AI, whereas less creative participants grew dependent on it. She also found that personality traits, such as conscientiousness, influenced how effectively users collaborated, with more conscientious individuals benefiting from working alone. Over time, users who worked alone showed continuous growth in creativity, whereas AI users did not exhibit similar gains. Although AI can enhance initial creativity, Luan [2024] concludes that human-only efforts can lead to more sustainable growth. She emphasizes the need for careful interface design for AI tools, suggesting adaptable AI that encourages user input and iterative exploration.

AI boosts initial creativity, but human-only users improve over time; social loafing and personality traits affect AI collaboration

2.4 Group Creativity

Creativity in groups is essential for solving complex problems and driving innovation. Unlike individual creativity, group creativity involves collaboration, idea sharing, and conflict resolution [Nijstad, 2015]. In his paper *Creativity in Groups*, Nijstad [2015] examines group creativity. He initially defines creativity as the production of novel, appropriate ideas, and contrasts individual creativity with group creativity. He continues by discussing two cognitive models: the matrix model [Brown et al., 1998], which frames idea generation as spreading activation, and the SIAM model [Nijstad and Stroebe, 2006], which shows how exposure to others' ideas can spark new ones. However, Nijstad and Stroebe [2006] note that group collaboration can also reduce creativity due to social dynamics, such as motivational loss, evaluation apprehension, and production blocking. Groups tend to be less productive at brainstorming than individuals, though written or typed brainstorming can reduce some blocking [Nijstad, 2015]. According to Nijstad [2015] group composition, diversity, and membership changes could predict group creativity. The key is to balancing creative and conformist members, which can enhance ideation and implementation. Also, task-related diversity often boosts creativity. Changing group members can also provide fresh perspectives, improving creative output [Nijstad, 2015].

Group creativity involves collaboration but can be hindered by social dynamics like motivational loss and production blocking

2.5 Generative AI in Group Creativity and Collaboration

Although there is limited research on group collaboration/creativity with AI, Kim et al. [2024] examine the role of ChatGPT in group brainstorming for creative idea generation in educational settings, in which a between-subject study was conducted on AI assisted and human-only groups. The findings show that AI-assisted ideas were rated as more creative than not AI-assisted ideas, and they also were rated as more detailed and diverse. Kim

ChatGPT enhances creativity in group brainstorming but risks user over-reliance

et al. [2024] stress that ChatGPT can enhance creativity in brainstorming by expanding the expression of ideas, but there is a risk of users relying too much on generative AI. They emphasize that further research is needed to balance the influence of AI on creative group collaboration and ensure diverse and original ideation.

LADICA supports team brainstorming and inclusivity but requires balanced AI-human interaction

Furthermore, Zhang et al. [2024] introduce LADICA, an AI-enhanced large shared display for co-located team collaboration in brainstorming, organizing and analyzing ideas. In their user study, participants used LADICA to brainstorm ideas, analyze and categorize them, and finally use as a group discussion reference. LADICA provides *cognitive scaffolding*, which is described as external cognitive support that helps group processes [Zhang et al., 2024]. The findings from this study showed that LADICA enhanced divergent thinking and inclusivity, however, Zhang et al. [2024] stress the need to balance the human-AI interaction to avoid over-reliance, and social loafing.

2.6 Ethical Considerations in AI

AI ethics require transparency, fairness, accountability, security, and user autonomy

It is also important to mention related work regarding the ethical considerations in AI. Balasubramaniam et al. [2022]'s *"Transparency and explainability of AI systems: From ethical guidelines to requirements"* emphasizes 5 attributes: transparency, privacy, fairness, accountability, and security. According to the article, AI systems must be transparent and explainable. This way users know how decisions are made and when AI is influencing outcomes. Also, privacy and data protection must follow regulations, with clarity around data anonymization. AI must be also trained on diverse datasets to prevent bias and discrimination. Furthermore, accountability needs human monitoring, and the AI must be secure and reliable, to prevent misuse and ensure consistent performance. Lastly, trustworthiness and user autonomy are essential. This way trust is achieved through explainable and ethical decision-making [Balasubramaniam et al., 2022].

Chapter 3

Research Questions

This chapter outlines the research questions developed for this thesis. The aim of conducting the user study is to explore the role of generative AI on co-located group creativity. Due to this being an exploratory study, the decision was made not to formulate a hypothesis. Therefore, the research remains exploratory with the intention of revealing unexpected findings. During both Study 1 and Study 2, we used a mixed methods approach to provide a combination of qualitative and quantitative results.

This research explores AI's role in group creativity using a mixed methods approach

According to Stadtländer [2009], qualitative research relies on exploring meaning, perspectives, and experiences, rather than numerical data. It is useful for examining social interactions and is often used in social sciences. The data used for qualitative research can include free-text questionnaires, interviews, and video recordings from the user study. Meanwhile, quantitative research is usually associated with experimental research, and it relies on numerical data and statistical tools. Quantitative data includes structured surveys and questionnaires. It is useful for generalizing findings from user studies and providing objective, measurable outcomes [Stadtländer, 2009].

Qualitative research explores experiences, while quantitative research provides measurable, generalizable data

We therefore chose a mixed-methods approach because it combines qualitative and quantitative methods, which allows us to capture both numerical trends and deep subjective insights in our research. This research benefits from

A mixed-methods approach enriches analysis by combining numerical trends with subjective insights

the mixed method approach, because it enhances both the depth and breadth of our analysis, and it allows us achieve triangulation by addressing our qualitative insights with our quantitative findings [Stadtländer, 2009].

3.1 Research Questions

RQs focus on generative AI's impact on group creativity

The study is structured around the following research questions (RQs), which explore the role of generative AI in creative group processes. We formulated a main research question, from which we developed a number of sub-questions.

RQ1 How do generative AI tools impact the creative process and outcomes in co-located group settings?

- **RQ1.1** How do AI tools influence collaboration dynamics in the creative process of the group?
- **RQ1.2** How do participants perceive the role of AI in creative group work?
- **RQ1.3** How does AI impact how group members perceive their group's creative result?
- **RQ1.4** How does AI impact how others perceive a group's creative result?

These research questions act as a framework for exploration, allowing our research to be analyzed qualitatively and quantitatively.

Chapter 4

Study 1: User Study

In this chapter, we present the foundation of our first study, named **Study 1**. We begin this chapter by outlining the preliminary study design that served as a guide for the pilot study, which was conducted to identify potential challenges and practical issues within the initial design. Afterwards, we provide a comprehensive overview of the final refined study design, which we shaped by the insights gained from the pilot phase. A sketch of the user study setup can be seen in Figure 4.1. In the next chapter, we describe Study 2, which is an online survey that allowed individuals to evaluate the creative results from Study 1.

4.1 Initial User Study Design

We learned from Nijstad [2015]'s *Creativity In Groups* that effective group dynamics can be fostered by encouraging all group members to participate and by minimizing fear of judgment. In addition, we discovered that we must structure the interactions in a way that reduces production blocking, which is when a participant's thoughts and ideas are blocked due to turn-taking. Therefore, in order to investigate the effects of artificial intelligence tools on creative processes in groups, we initially designed a user study, in which groups of three collaborate under three

Initial study design tests AI's impact on group creativity across three conditions

Figure 4.1: User Study Setup

different conditions without the need to take turns. Each group would have to complete creative tasks for each condition. The three conditions originally decided on were:

- No AI tools are used to perform the creative tasks.
- A generic AI tool, i.e. ChatGPT, should be used to perform the creative tasks.
- A custom AI tool, which provides limited outputs, should be used to perform the creative tasks.

The reason for deciding on these three conditions was inspired by the work of Noy and Zhang [2023] titled *Experimental evidence on the productivity effects of generative artificial intelligence*, in which the effects of AI on the productivity

of participants during writing were investigated. Their results showed an increase of productivity and enjoyment for users who used AI tools, and the largest increase of productivity was among users with weaker skills. Another study that inspired this decision is Magni et al. [2024]'s *Humans as Creativity Gatekeepers: Are We Biased Against AI Creativity?*, in which the authors investigate people's perception of the creative output of AI vs. humans. The findings of Magni et al. [2024]'s study show that people perceive AI outputs as less creative than human outputs, and people often find AI to exert less effort than humans when creating artifacts. The aforementioned papers inspired us to inspect the process and outcome of groups performing creative tasks while using AI tools and, in contrast, while not using AI tools.

Study conditions were inspired by research on AI's impact on productivity and biases against AI creativity

We also chose to focus our research on groups rather than individuals to address the **research gap** surrounding co-located groups using artificial intelligence tools in creative tasks. For the user study, we implemented a within-subject design, where all participants experience each condition of the independent variable. Our decision was informed by Charness et al. [2012]'s *Experimental Methods: Between-Subject and Within-Subject Design*. A within-subject design allowed us to directly compare participants' performance across conditions, making it particularly suitable for understanding how behavior and outcomes shift under different circumstances. Additionally, this approach required a smaller sample size, which was ideal given the relatively short and non-intensive nature of the study tasks.

A within-subject design was chosen to address the research gap in group AI creativity and allow direct condition comparisons

After deciding on our general study design concept, we made the decision to conduct a 60 minute study per group on average. This is in order to avoid participant fatigue, as explained by Schatz et al. [2012], while giving the participants enough time to complete their tasks and complete questionnaires.

Study duration set to 60 minutes to prevent participant fatigue

A challenge we faced was estimating the time required to complete a task during the user study, because groups tend to work at different paces. To stick to our planned 60-minute duration, we divided the 60 minutes into 5 minutes for a brief introduction, the Background Questionnaire

Task timing was structured to fit 60 minutes, with a pilot study ensuring feasibility

and the consent form. We then gave 15 minutes for condition's creative task, which were distributed into 10 minutes for the task and 5 minutes for the After-Task Questionnaire. The remaining 10 minutes were reserved for a short group interview to gather feedback on participants' experiences. To ensure that the allocated times were appropriate, we needed to conduct a pilot study before the main user study.

Marketing tasks were chosen for their visual, textual, and real-world relevance

When determining the nature of the creative tasks for the study, we aimed to select a task that naturally incorporated both visual and textual components, offering the potential for engaging results. Additionally, we looked for tasks that could be adjusted in terms of complexity and duration to suit the study's structure. Another important consideration was ensuring the task's relevance to real-world scenarios, making it both realistic and applicable. Based on these criteria, we decided that the tasks in the user study would center around **marketing** activities.

The marketing task was divided into sub-tasks

After deciding that the task objective was to create a marketing strategy for a specified item, we divided the marketing task into sub-tasks. The sub-tasks of the pilot study include (1) defining the target audience of the marketed item, (2) developing a catchy slogan, (3) drawing a logo for the product, and (4) highlighting the key selling points.

Products were chosen to spark creativity and avoid clichés

In addition, we formulated a selection of items to be marketed during the lab study, deliberately choosing unconventional products over typical ones. This approach was intended to stimulate imagination and foster collaboration among group members, while also minimizing clichés, therefore enhancing the diversity of responses across groups. The items that were chosen for the marketing task are:

- A Glow-in-the-dark Umbrella
- Edible cups
- Smart socks with fitness tracking

After deciding on the marketing products, we decided to counterbalance the three conditions and three products. Counterbalancing, according to DePuy and Berger [2014], is a technique that is used to minimize the influence of order effects. The technique involves varying the sequence in which treatments or tasks are presented to participants.

Counterbalancing was used to reduce order effects in the study

Regarding the group size, it was decided to form groups consisting of three members, which is the minimum number needed to fulfill the group requirement.

Groups were limited to three members

4.1.1 Custom AI Preparation

Our idea behind the custom AI condition was to create a custom GPT model, which generates limited outputs compared to the normal ChatGPT. The custom GPT, named Creative Companion, was trained to only generate textual responses that guide the users with their tasks, rather than solving their tasks completely. It was designed to only provide textual responses, even for visual tasks. For example, when a user would ask for help with a visual task, such as drawing a logo, the AI would respond with ideas to implement the idea, such as possible themes, symbols, and font ideas. An example of this type of interaction can be seen in Figure 4.2.

Creative Companion was designed to guide users with limited textual outputs, not solve tasks

The custom AI was intentionally designed to ensure that its outputs were less polished than those typically produced by the users themselves. Its primary function was to provide inspiration and guidance, offering rough ideas and directional suggestions rather than fully developed solutions. This design choice aimed to explore how a limited AI impacts the creative process within groups, particularly when compared to standard AI or the absence of AI. The focus was on examining the role of AI as a supportive tool rather than as a dominant contributor to the creative workflow.

Custom AI provided rough ideas to inspire, not replace, user creativity

Furthermore, the initial study designs included a plan for the data collection. The data that would be collected includes:

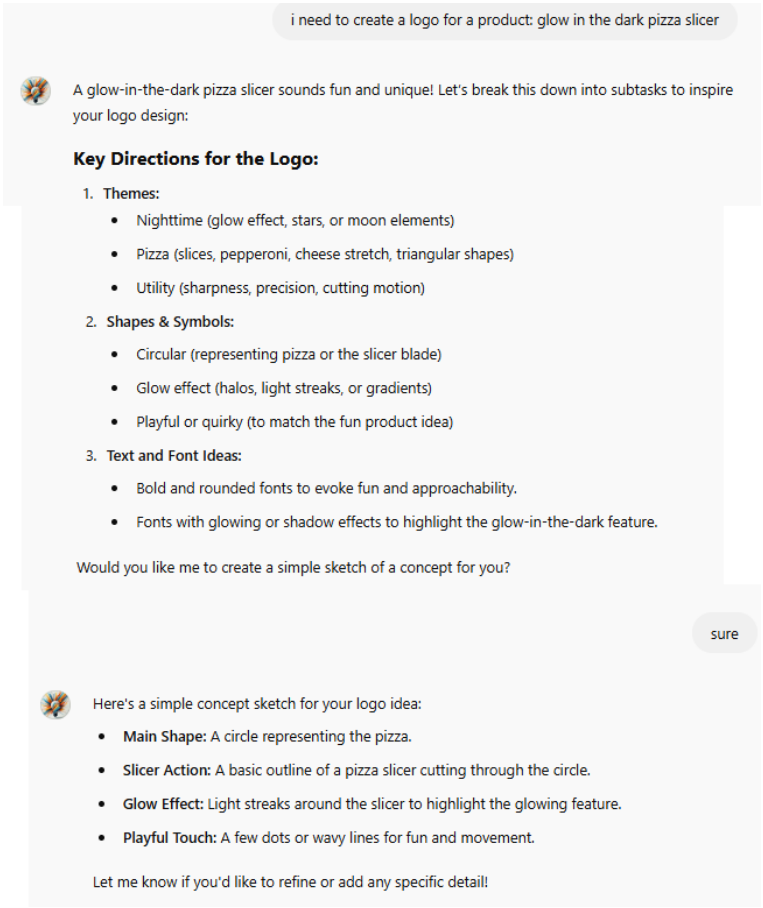


Figure 4.2: Custom AI Interaction Example

- **Before Any Tasks:**
  - Demographic Information
  - Self-perceived Creativity of Individuals
  - Comfort with Group Work
- **During Each Task:**
  - AI Interaction Logs
  - Collaboration Dynamics
  - Task Outputs
- **After Each Task:**
  - Collaboration Dynamics

  - Task-Specific Creativity Perception
  - Emotional and Cognitive Load
  - AI's Role and Perceived Usefulness (if applicable)
  - Open-Ended Feedback

- **After All Tasks:**
  - Comparative Questions
  - Suggestions and Improvements
  - General Insights

The final part of the study design addresses ethical considerations. We used Yu [2020]'s *Ethical Considerations in Case Studies* as guidance for our ethical considerations. His paper resulted in 21 ethical guidelines for conducting user studies. We followed the guidelines when applicable, with the goal of maintaining a fair and ethical user study. The main guidelines we followed include *informed consent*, *voluntary participation*, *avoiding conflict of interest*, and *transparency in study objectives*. We also ensured the privacy and confidentiality of data, such as the video recordings and documents from the study. We also tried to minimize participant discomfort, respect cultural sensitivities, and ensure data security.

Ethical guidelines ensured consent, transparency, privacy, and participant well-being

4.2 Pilot User Study

After the initial study design was ready, the pilot study was conducted on three volunteers, of whom two participants have experience in pilot studies and user testing. Presser et al. [2004]'s *Methods for Testing and Evaluating Survey Questionnaires* was used as a reference during the preparation of the pilot study, and for creating the meta feedback sheet that was provided exclusively in the pilot study. Presser et al. [2004] suggest combining multiple testing methods during a pilot study, because using various methods can help identify different types of problems during the pilot study. This can be achieved through interviews, feedback

The pilot study used multiple evaluation methods to identify potential issues

sheets and coding the behaviors of the participants [Presser et al., 2004].

4.2.1 Preparation

A large screen and wireless input devices ensured seamless collaboration

For the pilot study, a very large screen was acquired, which was connected to the main computer. The participants used the same computer and screen during the two AI conditions. Additionally, a wireless keyboard and mouse were provided to ensure that participants could type comfortably and take turns seamlessly, without the need to physically switch seats.

Participants received forms, provided feedback, and took part in a semi-structured interview

The participants were seated in chairs facing the screen, and were given blue pens. They were also handed out documents at different stages of the study, starting with the consent forms, then the demographic forms, and a questionnaire after each task. Since this was a pilot study, they were also given a meta feedback sheet, where they could fill out feedback regarding how the study was conducted. Finally, the participants were interviewed in a semi-structured interview, in which they were asked for feedback regarding the conditions and tasks, as well as feedback on the study and issues that arose during it.

The pilot study was filmed with an external HD webcam, that was attached to the same computer on which the study was conducted. The duration of the pilot study was 2 hours and 10 minutes.

4.2.2 Meta Results

Pilot study feedback informed refinements to the final study design

During the pilot study, participants were given a meta feedback sheet to provide comments on the details of the study itself. The meta feedback sheet can be found in Appendix A. The pilot study resulted in new decisions for the finalized user study design, which are marked in **bold**.

Questionnaires were revised for consistency based on participant feedback

All participants reported that the instructions provided by

the researcher at the beginning of the study were clear. However, they found some aspects of the questionnaire sheets poorly structured and unclear. Specifically, they noted inconsistencies in the scaling, with some questions requiring ratings between 1 and 5, while others used a scale of 1 to 20. **Based on this feedback, the questionnaires were revised to use a universal scale of 1 to 5**.

When asked about the ease of completing the study tasks, two participants rated it as neutral, while one found it easy. Additionally, two participants felt that the tasks were applicable to real-life scenarios. However, one participant pointed out that in a real marketing team, the product to market would already be well-defined, unlike the tasks in the study.

Participants found tasks realistic but noted differences from real marketing teams

All three participants agreed that the allocated time was insufficient for completing the tasks. Although they were initially informed that the tasks would take approximately 10 minutes, they required 15–20 minutes per task. **As a result, the task duration was increased to 20 minutes for the user study**. Additionally, **it was decided not to explicitly tell participants how much time they should take to finish a task**, allowing them to work at their own pace. Also, none of the participants reported feeling that the researcher's behavior or comments influenced their task performance. Regarding data collection, one participant expressed discomfort with having their face recorded. **Consequently, the decision was made to film participants from behind during the user study to address privacy concerns**.

Task duration was increased, time limits removed, and filming adjusted for privacy

During the pilot study, participants were asked to work in English. Two of the three participants suggested that they would have preferred to speak in their native language, German, during the tasks. In response, **it was decided that participants could speak to one another in their mother tongue**, provided the researcher also understands the language. However, the task sheets would still need to be completed in English. Furthermore, an optional feedback section revealed that one participant enjoyed using AI to generate ideas, while another appreciated the first two tasks, both of which involved AI. Participants also requested clearer instructions for the tasks. Regarding

Participants allowed to speak in their native language, but task sheets remained in English

the products being marketed, participants found the edible cups to be too broad as a concept and the smart socks with fitness tracking to be too specific. In contrast, they felt that the Glow-in-the-dark Umbrella struck the perfect balance—specific enough to understand the product, yet vague enough to allow room for creative marketing ideas.

4.2.3 Pilot Study Results

Over the course of the pilot study, the participants worked under a condition without AI and two conditions with AI: one using ChatGPT and the other using our custom GPT, *Creative Companion*. The custom AI was specifically designed to provide limited outputs, such as offering general ideas and guidance to help solve the tasks, rather than producing detailed results. However, the custom AI did not function as intended and occasionally generated more detailed responses than expected.

The custom AI sometimes provided more detailed responses than intended

For instance, when participants initially asked for a slogan, the AI provided keywords as intended. However, when they followed up with a request for a slogan containing a specific word, the custom AI generated multiple fully-formed slogan options, contrary to its intended behavior. Similarly, the custom AI was designed to describe images rather than generate them. Despite this, the participants' persistent requests for an image eventually bypassed the programmed restrictions, resulting in the AI generating images, which it was not supposed to do.

Custom AI occasionally bypassed restrictions, generating full slogans and images

The issues with the custom AI not functioning as intended, combined with the fact that participants required closer to 20 minutes rather than 10 minutes per condition, led to **the decision to completely remove the custom AI condition from the study**. This decision addressed the time constraint, reduced the overall workload for participants, and eliminated the complications caused by the custom AI's unexpected behavior. This led to a significant change in the study design, reducing the number of conditions from three to two. As a result, the final study would focus solely on comparing two conditions: using a standard AI tool (Chat-

The custom AI condition was removed to address time constraints and AI inconsistencies

GPT) versus not using any AI tool to complete creative tasks in a group.

In addition, during the pilot study, participants received information about the item they were marketing on a board, which displayed the name of the item and the associated sub-tasks. They were also provided with blank paper and standard blue pens to complete their tasks. Observations from the pilot study led to the **decision to use structured worksheets** with designated spaces for each sub-task, creating a more organized experience for both participants and researchers. Additionally, it was decided to **replace the basic blue pens** with higher-quality, colorful pens to encourage a more comfortable and engaging environment, which could potentially boost the participants' performance. We also decided to provide each participant with **an individual handout** containing the name and information about the product being marketed, along with the sub-tasks, instead of using a single large board, with the goal of increasing comfort.

Structured worksheets, colorful pens, and individual handouts were introduced for better engagement

4.3 Finalized User Study Design

The finalized study design involves a user study conducted with groups of three participants, who collaborate on marketing specified products. Each group needs to complete the marketing task twice: once while using a generative AI tool (ChatGPT-4o), and once without AI-assistance. Before beginning the study, participants have to sign a consent form. Afterwards, the researcher conducting the study would give an introduction to the task. Participants are asked if they had any time restrictions, since the study is designed to take approximately 60 minutes, but this can vary depending on the group's pace. Then the researcher asks the participants to fill out a demographic questionnaire, and soon after outlines the structure of the study. The participants are told that they will work on tasks in two conditions, the AI and No AI conditions. Also, the researcher would mention the marketing task and its sub-tasks, and the fact that the participants have to fill out multiple questionnaires. Additionally, participants are informed that the

Final study design includes two conditions, clear instructions, and privacy-conscious video recording

study is video recorded from behind to ensure privacy and that all collected data is securely stored and deleted upon thesis completion. The participants are also given the opportunity to refuse participating in the study if they are uncomfortable with the video recording or other aspects of the study. Following this introduction, the researcher distributes a handout detailing the item to be marketed and the condition in which the participants would work. This way, the participants are provided with the necessary context to begin their task.

The pilot study resulted in a major adjustment to the finalized study design. The most significant change was the removal of the custom AI condition, simplifying the study to two conditions: (1) **AI Condition** (ChatGPT), and (2) **No AI Condition**. There was also a slight change made to the No AI condition after the pilot study, in which participants became allowed to use Google, with the limitation of not using any generative AI tools. The changes applied to the study design after the pilot study also lead to an increase in the time allocated for each condition, extending the duration of each main task from the originally planned 10 minutes into 20 minutes. Despite this adjustment, the overall study duration remained at approximately 60-65 minutes. The revised schedule includes 5–10 minutes for the introduction, consent form, and demographic questionnaire, followed by 20 minutes for the marketing task and 5 minutes for the post-task questionnaire, for each condition. The final phase, lasting about 5 minutes, is dedicated to the brief Final Questionnaire, which would capture opinions on both conditions.

Study refined to two conditions, with extended task duration and limited Google use in No AI condition

In terms of language use during the study, English is the language chosen for task outputs. However, participants are permitted to communicate with each other in their native language, provided the researcher also speaks that language. The decision to conduct the study in English is based on the requirements of the second phase, where the creative results of Study 1 are evaluated and reviewed. Ensuring a consistent language across all outputs helps maintain clarity and avoids possible confusion during Study 2.

Study outputs are in English, but participants may communicate in their native language

4.3.1 User Study: Marketing Task

During each condition, the participants would work on marketing a specific product. They received information about the object they are marketing on a handout, which includes the name of the product, minimal information about the product, and also the list of the four sub-tasks that the participants are expected to complete.

Participants receive a handout with product details and task instructions

Marketing was selected as the main task for the user study for several reasons. First, it is a realistic scenario where groups commonly collaborate on creative tasks composed of multiple sub-tasks. This makes it an ideal choice for investigating group dynamics and creativity. In terms of the products chosen for the marketing tasks, the first item, the Glow-in-the-dark Umbrella, was selected prior to the pilot study. As outlined in section 4.1, the selection criteria emphasized items that were unusual and creative, with the purpose of fostering creativity and reducing the possibility of clichés. After the pilot study, adjustments were made to the second product, leading to the selection of the Smart Plant Pot. As a result, the final products chosen for the study are:

Marketing was chosen for its realistic, collaborative, and creativity-driven nature

1. Glow-in-the-dark Umbrella
2. Smart Plant Pot

These choices were designed to create a balance between creativity and practicality. In addition, the sub-tasks were revised for the marketing task. The sub-tasks include: (1) naming the product, (2) creating a slogan, (3) listing 2-3 selling points, and (4) drawing the product with a logo.

Handouts

During the pilot study, the participants are shown the information about their task on a board, which included the name of the product and the name of each sub-task they were expected to do. However, it was decided to provide

handouts with this information for each participant to provide them with more comfort and more information about the sub-tasks, including examples.

Handouts replaced the board for clarity, comfort, and structured task guidance

A handout is given to each participant for each product. Each handout provides a structure for participants to develop the marketing strategy for the specific product. It includes a clear task description, sub-tasks broken down into four steps, which are (1) naming the product, (2) creating a catchy slogan, (3) listing 2-3 key selling points, and (4) drawing a the product and its logo. Additional notes clarify the materials provided, the importance of using the worksheet for writing their answers, and the importance of group collaboration during the task. In addition, a solved example worksheet is provided for the participants to view for 60 seconds, in which the same sub-tasks are solved for a different product, specifically an *edible cup*. Before the participants start working on their task, the example sheet is taken away in order to avoid creating bias. The handout can be seen in Appendix A, titled *Marketing Task Instructions*.

Worksheet

Structured worksheets ensure clarity, consistency, and reliable data collection

Following the pilot study, it was decided to provide participants with a structured worksheet. This decision was made to ensure that groups complete all sub-tasks and record their answers clearly, particularly since their responses are written on paper. Structured worksheets are essential not only for the researcher to accurately collect and organize data but also for generating consistent outputs. These outputs are evaluated afterwards in Study 2, where participants complete an online survey assessing the creative results produced in Study 1.

4.3.2 Counterbalancing

After revising the structure of the user study to remove the custom AI condition, the need for counterbalancing across

three conditions and three marketing products was eliminated. The finalized study required counterbalancing only two conditions, **No AI** and **AI**, along with two products, referred to as **U** (Umbrella) and **P** (Plant Pot). The updated counterbalanced structure is as follows:

Counterbalancing adjusted for two conditions and two marketing products

- Group 1: No AI + **U** → AI + **P**
- Group 2: No AI + **P** → AI + **U**
- Group 3: AI + **U** → No AI + **P**
- Group 4: AI + **P** → No AI + **U**
- Group 5: No AI + **P** → AI + **U**
- Group 6: AI + **U** → No AI + **P**

Counterbalancing is essential to ensure that groups experience both conditions (AI and No AI) in different orders. This effectively addresses potential order effects and minimizes the risk of sequence-related biases in the final results. Also, alternating which product is paired with each condition can help prevent product bias, meaning that no particular condition is always associated with a specific product DePuy and Berger [2014].

Counterbalancing prevents order effects and product-related biases

4.3.3 Questionnaires

Regarding the questionnaires, each participant is given an individual form to complete independently. The first questionnaire is the Background Questionnaire, which is divided into four sections. Many questions from Background Questionnaire were derived from or inspired by the article *Demographic survey questions that yield valuable insights*. [1]

Each participant completes an individual Background Questionnaire

[1] `https://www.qualtrics.com/blog/demographic-survey-questions/`

Questionnaire 1: Background Questionnaire

The **first section**, titled General Information, has basic questions about age, gender, education, and profession. In *Creativity Background*, participants answer questions on their experience with creative activities. They are asked how often they work on creative activities and how they rate their own creative abilities. In addition, they are asked if they have experience working on a creative group project. In *Familiarity with AI*, participants are asked about their experience with AI tools. This includes questions on whether they have used AI tools, how frequently they use them and why they used them. Moreover, they are also asked to assess their familiarity with these AI tools. The final section, *Group Work Preferences*, explores participants' opinions and attitudes toward teamwork. They are asked to rate how much they enjoy working in a group and to indicate what role they prefer to have in a group. The full questionnaire is included in Appendix A for reference.

The Background Questionnaire covers demographics, creativity, AI experience, and group work preferences

Questionnaire 2: After-Each-Task Questionnaire *(Both Conditions)*

The second questionnaire, titled the After-Each-Task Questionnaire, is available in **two versions**: one for the AI condition and another for the No AI condition. Both versions are organized into five sections: Collaboration Dynamics, Task-Specific Creativity Perception, Emotional and Cognitive Load, AI Tool's Role or Lack of AI Tool's Role, and Open-Ended Feedback. The questionnaire has single-choice tables for the first four sections. They all use a five-point scale that varies depending on the section. For example, the *Collaboration Dynamics* section uses a scale ranging from "*Strongly Disagree*" to "*Strongly Agree*". In addition, each question is numbered based on its section and position within that section. The aim behind this is to enhance clarity for both the researcher and the user study participants. For example, the second question in the first section is labeled as 1.2. The questions for this section of the questionnaire were adapted from the Team Effectiveness

The After-Each-Task Questionnaire assesses collaboration, creativity, cognitive load, and AI's role

	-2	-1	0	1	2	
2.1. Conventional	O	O	O	O	O	Original
2.2. Predictable	O	O	O	O	O	Unexpected
2.3. Unfocused	O	O	O	O	O	Well-defined
2.4. Impractical	O	O	O	O	O	Practical
2.5. Unappealing	O	O	O	O	O	Appealing
2.6. Disorganized	O	O	O	O	O	Organized

Figure 4.3: Example from After-Task-Questionnaire: Task-Specific Creativity Perception Questionnaire

Questionnaire developed by the University of Colorado[2]. Only four questions were selected from the original questionnaire, based on their relevance to our research.

In the second section of the questionnaire, *Task-Specific Creativity Perception*, participants have to evaluate their group's marketing result using a scale ranging from -2 to 2. Each row presents a pair of opposing adjectives (e.g., "*Conventional*" vs. "*Original*"), and participants must to select one number for each adjective pair. The purpose of this section is to evaluate how participants perceive their creative results in both conditions.

Task-Specific Creativity Perception measures how participants rate their group's creative output

This section of the questionnaire also enables direct comparisons of ratings within a team and provides a foundation for comparing internal perceptions of creativity with evaluations from external raters, which are collected during Study 2. The questions in this section are adapted from the Creative Product Semantic Scale (CPSS), a scale developed by Besemer and O'Quin [1986] to measure the creativity of a product. However, the original CPSS, which includes 15 adjective pairs divided into three categories, novelty, resolution, and style, was reduced to six pairs in this study to eliminate redundancy and overlap in meanings. The Task-Specific Creativity Perception questions can be seen in Figure 4.3.

[2] `https://www.cu.edu/sites/default/files/Team_effectiveness_questionnaire.pdf`

The third section of the questionnaire is a modified version of the NASA Task Load Index (NASA-TLX) [Hart, 2006], which is a tool that is commonly used for assessing participants' perceived workload during a task. It evaluates six factors, mental demand, physical demand, temporal demand, performance, effort, and frustration level. The questionnaire can offer insights into task difficulty and overall participant experience [Hart, 2006]. For our user study, the physical demand and frustration factors were excluded, as they are not relevant to the scope of the study. The remaining four factors were included to compare variations in workload, such as mental demand and effort, between conditions where participants use AI and those where they complete tasks without AI, highlighting the impact of AI tools on group dynamics and task performance.

A modified NASA-TLX assesses workload differences between AI and No AI conditions

The final scale-based section of the After-Task Questionnaire has two versions. One version is for the AI condition, named *Digital Tool's Role and Perceived Usefulness*. In this questionnaire, we purposely avoid using the words AI or Artificial Intelligence, in order to reduce pre-existing bias. Instead, we use the expression *Digital Tool*. For this section we use a scale ranging from Strongly Disagree to Strongly Agree. The questionnaire for the AI condition is named *Digital Tool's Role and Perceived Usefulness*, and the questionnaire for the No AI condition is called *Lack of Digital Tool*. Both versions of the this questionnaire section have the same first 7 questions, while the AI-version has 3 extra questions. The questionnaire is listed below.

AI was referred to as a "Digital Tool" in the questionnaire

After Task Questionnaire - Q4 - **Both Conditions**:

- I felt inspired.
  - Reason: This question assesses the condition's impact on participants' creativity and motivation. It provides insight into whether the tool or lack of it influenced their inspiration during the task.
- The task was enjoyable.
  - Reason: This measures the enjoyment of participants with the task itself, regardless of the con-

dition. Assessing this across both conditions can help us determine if the use of AI or the absence of it influences participants' general satisfaction with the task.

- Our team collaborated well.
  - Reason: This measures the effect of the condition on group dynamics and collaboration, exploring whether the tool or absence of it enhanced or hindered teamwork.
- I appreciated the creative input of my team.
  - Reason: This measures how much participants valued and relied on their teammates' contributions during the task. This highlights the role of group collaboration in the absence of AI support.
- I needed more help.
  - Reason: This evaluates the participants' perceived adequacy of the condition in supporting their problem-solving needs, highlighting areas where additional assistance might have been required.
- Our team worked too slowly on our task.
  - Reason: This assesses the condition's impact on task efficiency, identifying whether the presence or absence of the tool contributed to delays or inefficiencies in task completion.
- I could think independently.
  - Reason: This question evaluates the extent to which participants felt empowered to generate ideas and solve problems autonomously without relying on external tools, in this case AI.

After Task Questionnaire - Q4 - **AI condition**:

- I felt distracted by the digital tool.

  - Reason: This question evaluates whether the AI tool diverted attention from the task, providing insights into its potential to hinder focus during group work.

- I was confident about when to use the digital tool.
  - Reason: This assesses participants' understanding of the tool's functionality and their confidence in deciding when and how to use it effectively during the task.
- The digital tool limited my creativity.
  - Reason: This explores whether participants felt the AI constrained their creative process, highlighting potential negative impacts of using AI tools on ideas generated by the group.

The fourth section enables detailed AI vs. No AI comparisons

The fourth section of After-Task Questionnaire was therefore designed to gather comprehensive insights into the two different conditions and enable a detailed comparison of specific aspects between them.

Finally, the fifth section of the After-Task Questionnaire is the open-ended feedback section. Both the AI and No AI version have the same three questions, and the AI condition has an extra question. To avoid confusion, there is a disclaimer stating that *This Condition = Using the Digital Tool* **or** *This Condition = Not using the Digital Tool* depending on the condition. The expression "AI" is also avoided here to avoid potential bias. Thus, the questions asked in the open-ended section are:

The final section gathers open-ended feedback on each condition's benefits and challenges

- What were your overall thoughts or impressions about working in this condition?
  - Reason: This question provides insight into participants' general perceptions toward the condition, offering valuable qualitative feedback about their overall experience.
- What aspects of this condition did you find most helpful, if any?

  - Reason: This question helps identify specific elements or features of the condition that participants perceived as beneficial.

- What challenges did you face while working under this condition, if any?
  - Reason: This question helps us identify the difficulties that participants encountered. This helps us understand the limitations of the condition.
- *(AI Condition Only)* In what ways did you use the digital tool, and for what purpose?
  - Reason: This question explores how participants engaged with the digital tool. It reveals their patterns when using AI.

This section of the questionnaire provides the participants with a blank box to write their answers.

Questionnaire 3: Final

The Final Questionnaire is distributed to participants after they complete the task for the second condition. The purpose of this questionnaire is to gather unbiased feedback from each participant individually, in order to avoid the risk of uniformity of answers. The complete questionnaire can be found in Appendix A.

The Final Questionnaire collects individual feedback to prevent response bias

This questionnaire consists of 6 questions that were designed to gather participants' general impressions and preferences regarding the study. In the **first question**, the participants are asked which product they preferred marketing. The main purpose of this question is to enable participants to distinguish their evaluation of the condition from their opinion of the product. Building on this, the **second question** shifts the focus to their preferred condition (using AI vs. not using AI), which provides insight into how participants perceived the tools' influence on their task experience.

Questions distinguish product preference from condition preference

Questions assess AI's impact on creativity, collaboration, and perceived role in teamwork

The **third question** assesses which condition the participants felt produced the most creative results, helping identify the influence of AI on creativity. Similarly, the **fourth question** asks about group collaboration, specifically under which condition the team worked best together, which provides insight into the dynamics of group interaction. Moreover, the **fifth question** asks whether the AI feel more like a teammate, a guide or a leader during the tasks. The first five questions are all single-choice questions, with the answer options listed below them along with check-boxes.

The final question allows open-ended feedback on the study experience

Lastly, the **sixth question** gives participants the option to reflect on their overall user study experience by offering qualitative feedback. This may reveal perspectives that are not addressed in the structured questions. The sixth question is optional, and provides the participant with an empty box to write their answer.

Final questions ensure individual, unbiased participant feedback

The aforementioned questions are designed to gather participants' individual and unbiased opinions about their experiences and preferences across the study conditions. This allows us to capture their perspectives individually, avoiding the risk of bias or uniformity that might occur in a group interview.

4.3.4 Summary of Study 1 Procedure

Study 1 summary

To summarize, Study 1 is a 60-minute within-subject user study designed to explore the impact of generative AI tools on creative co-located group processes. Each session involves a group of three participants and begins with a brief introduction to the study's concept, followed by participants each signing a consent form. This form informs them about their rights as participants, the video recording of the session, and the anonymization and disposal of all data after the thesis is completed.

Participants then separately fill out a Background Questionnaire that gathers demographic information, creativity levels, familiarity with AI, and group work preferences. This initial phase takes approximately 5–10 minutes. Af-

terward, we collect the questionnaires and introduce the first task, explaining the assigned condition (AI or No AI), the marketing task structure, and the sub-tasks. During the No AI condition, participants are given the option to use Google to browse for information when needed, with the single restriction of not using artificial intelligence tools. During each condition, participants are told which product they will market, either a Glow-in-the-dark Umbrella or a Smart Plant Pot, and are each given a handout detailing the task, sub-tasks, and product name. The handout is shown in Appendix A under the name *Marketing Task Instructions*.

Study 1 summary continuation

To provide clarity without restricting creativity, participants are briefly shown an example marketing sheet completed for a different product, an *edible cup*. After the participants view it for one minute, the example sheet is removed to avoid biasing their creative process. This example sheet can be found in Appendix A under the title *Marketing Task Example*. Participants are then provided with a single group worksheet containing the sub-tasks and spaces for their answers, as well as individual sketching sheets for brainstorming and testing ideas. They are instructed to use the sketching sheets freely but to record their final answers only on the official worksheet. Exactly one worksheet is submitted per group. The worksheet can also be found in Appendix A under *Marketing Worksheet*.

One task takes approximately 20 minutes to complete, after which participants separately fill out an After-Task Questionnaire that takes about 5 minutes. This questionnaire consists of five sections: Collaboration Dynamics, Task-Specific Creativity Perception, Emotional and Cognitive Load, Digital Tool's Role and Perceived Usefulness (or Lack of Digital Tool), and open-ended feedback. The same process is repeated for the second condition using the other product.

Each task lasts 20 minutes, followed by a 5-minute questionnaire

Conditions and products are counterbalanced to ensure fairness and minimize potential biases among the groups. Once both conditions and their respective questionnaires are completed, participants separately fill out the Final Questionnaire. This final step gathers insights about their preferred condition and product, their impressions of the

Study 1 ends with a Final Questionnaire on preferences and AI impressions

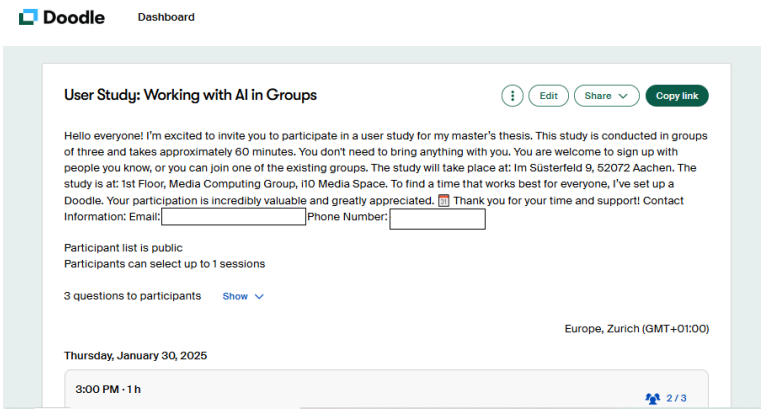


Figure 4.4: Doodle Scheduling Tool

AI, and overall feedback about the study. This Final Questionnaire takes approximately 5 minutes to complete, marking the end of the user study session.

4.4 Acquisition of Participants

Participant scheduling was managed via a Doodle poll

For the first study, we aimed to form groups consisting of exactly three participants. To organize the sessions, a Doodle poll with various time slots was created. This allowed participants to each select a group with their preferred time-slot. The invitations were primarily shared through local university groups.

The Doodle poll included three questions for participants

- Question 1: Do you have experience using generative AI tools (i.e. ChatGPT, DALL-E, etc.)?
  - Reason: Ensures that each group includes at least one participant familiar with using generative AI tools, to achieve balanced group dynamics and meaningful data collection.
- Question 2: Are you comfortable speaking English?
  - Reason: As the thesis and user study are conducted in English, it is essential for participants to have a sufficient level of English proficiency.

- Question 3 *(Optional)*: If your group cannot find a suitable time slot, what is your preferred time?
  - Reason: Allows flexibility for participants whose schedules do not align with the provided time slots, ensuring greater inclusivity and participation in the study.

Finally, when all 6 groups had registered on Doodle, the sign-up sheet was closed and the contact data of the participants was temporarily saved to send groups reminders before their respective appointments.

4.5 Results: Study 1

In this section, we present the results from Study 1, specifically the questionnaire results.

4.5.1 User Study Numbering System

In preparation for conducting the user study and evaluating its results, a numbering system was created with the goal of organizing the documents based on the group, participant, document type, and other details. The numbering system is explained briefly below, and can also be found in Appendix A.

A numbering system was created to organize study documents efficiently

- Study ID: Each group has a Study ID ranging from **01** to **06**.
- Participant ID: Each participant has a Participant ID within their own group based on their seating position, ranging from **X** to **Z**.
- Condition Type: The AI condition is labeled as **A**, while the No AI condition is labeled as **N**.
- Product Type: The products used in the marketing task are labeled based on their first letter, therefore

the umbrella is given the label **U**, while the plant pot is labeled as **P**.

- Document Abbreviations: Each document is given its own label or abbreviation to ensure clarity for reference and analysis. The label for each document is listed in Table 4.5.
- Numbering Order: In Table 4.6, the numbering order used to generate a document name can be seen. The number of a specific question can be attached to the end of its document name.

Label	Document Name
C	Consent Form
B	Background Questionnaire
H	Handout
W	Worksheet
S	Scribble Sheet
Qa	After Task Questionnaire (AI Condition)
Qn	After Task Questionnaire (No AI Condition)
F	Final Questionnaire

Table 4.5: Table: Document Abbreviations

Group	User*	Document	Condition*	Product*	Question*

Table 4.6: Table: Numbering Order

* If applicable.

4.5.2 Background Questionnaire Results

Demographic Results

The study had 18 participants, balanced by gender, with a strong STEM representation

The user study had 18 participants (9 male, 9 female), with a mean age of 25.6 years (min = 20, max = 29, SD = 2.7). Regarding their education, the majority have completed/are pursuing a Master's degree (8), followed by 7 Bachelor's degree students/graduates, and 2 Medical State

Exam students/graduates, as well as one high school graduate. The highest level of education reported represents the participants' current program or their completed degree. When asked about their fields of study/work, Computer Science (8) is the most common answer. Other fields include Medicine, as well as Psychology, and five additional scientific or engineering fields, each represented by one participant. The sample was balanced regarding gender, with a high representation in STEM fields.

Creativity Background

The Creativity Background section first addresses the level to which participants engage in creative activities. The most common answer is once a month (8), followed by 2-3 times a week (4) and once a week (4), and never (2). In regard to group collaboration on creative activities, half of the participants reported no experience, while the other half indicated engaging in different creative group contexts, such as architectural projects, crafts, jam sessions and more. Regarding the participants' self-perceived creativity, the mean rating was 2.93 on a scale of 1 to 5 (min = 2, max = 4). Participants had varied but moderate creative backgrounds, with mixed levels of experience with group creativity.

Participants had varied creative backgrounds, with mixed experience in group creativity

Familiarity with AI

The Familiarity with AI section reveals that all 18 participants have used generative AI tools before. ChatGPT is the most widely used by all participants, followed by Midjourney, Claude-son, DALL·E and other. With regards to the frequency of usage, 7 participants use the AI tools daily, 5 use them weekly, and the remaining 6 use them rarely or monthly. The primary reasons for using AI tools vary, with educational purposes being the most common (14 mentions), followed by work-related tasks (13), personal projects (12), entertainment (7), and writing emails (1). In measuring their familiarity with AI tools, the participants rated on the scale of 1 to 5 with a mean of 3.28 (min = 2, max

All participants had prior AI experience, with varying frequency and familiarity levels

= 5). The results show that groups were generally comfortable with AI tools, though skill levels and usage patterns differed.

Group Work

In this section of the Background Questionnaire, the participants provide insight into their familiarity with their group members members and their preferences regarding group work. Participants rated their familiarity with their group members with a mean of 2.4 (min = 1, max = 5). They participants also rated their enjoyment for group work with a mean of 3.11 (min = 2, max = 4) on a scale of 1 to 5, showing that most participants like working in a group to some degree. Regarding their preferred role in group work, the most common answer was equal contribution (9 participants), and 5 participants had no preference, with 3 participants preferring leading roles, and only one participant preferring secondary roles. This points towards a generally collaborative mindset, in which most participants preferred shared responsibility over hierarchical team dynamics.

Participants showed a preference for equal contribution in group work, with moderate enjoyment and varied familiarity

4.5.3 After Task Questionnaire Results (AI and No AI)

Since this study used a within-subjects design, meaning that the groups completed their task in both the AI and No AI conditions, the standard Cohen's d for independent groups was not the appropriate method. This is especially due to the small number participants in this user study. Instead, we used **Cohen's** d_z (paired-samples Cohen's d). This method accounts for the fact that each participant served as their own control. Therefore, this measure provides a better estimate of effect size by accounting for the variability of differences within individuals, rather than between groups [Goulet-Pelletier and Cousineau, 2018].

Cohen's d_z was used to account for within-subject design and individual variability

The reason behind using Cohen's d_z is due to three reasons. The first reason is that it accounts for within-subject

Cohen's d_z was chosen for its ability to account for within-subject variability and paired samples

variability, since the same participants provided ratings for both AI and No AI conditions, and this method reduces the influence of individual differences. The second reason is that Cohen's d_z corrects for paired samples. In a within-subjects design, responses are not independent. This means that the difference between conditions is more meaningful than individual condition means. The third reason is that Cohen's d_z is a statistically stronger method than Cohen's d in this particular case. This is because Cohen's d_z is more sensitive to detecting effects, in comparison to Cohen's d. This is because it removes between-subject noise [Goulet-Pelletier and Cousineau, 2018].

Calculation Method

For each questionnaire item, Cohen's d_z was calculated using the following steps:

Initially, for each participant, the difference between the AI condition score and No AI condition score was calculated:

$$D_i = X_{\text{AI},i} - X_{\text{No AI},i}$$

where:

- D_i = difference score for participant i
- $X_{\text{AI},i}$ = participant i's response in the AI condition
- $X_{\text{No AI},i}$ = participant i's response in the No AI condition

Afterwards, the mean of all difference scores was computed:

$$\bar{D} = \frac{\sum D_i}{N}$$

where:

- $\bar{D}$ = average difference score
- N = number of participants

Subsequently, the standard deviation of the difference scores was calculated:

$$s_D = \sqrt{\frac{\sum(D_i - \bar{D})^2}{N - 1}}$$

where:

- s_D = standard deviation of difference scores
- $N-1$ is used for sample standard deviation correction

Finally, Cohen's d_z was calculated as:

$$d_z = \frac{\bar{D}}{s_D}$$

This standardizes the mean difference in terms of the variability of the within-subject differences.

Cohen's d_z was calculated by finding the mean difference between AI and No AI ratings for each participant

We applied the calculations to the questionnaire data from our user study by recording each participant's rating for each question, once for the AI condition and once for the No AI condition. Then, for each pair of ratings on the same question, the difference was calculated by **subtracting the no AI score from the AI score**. This resulted in a table of (within-participant) difference scores for each question. Afterwards, the mean difference was calculated by averaging these differences that were calculated across all 18 participants for each question. In addition, the standard deviation of differences was calculated for the same set of values [Goulet-Pelletier and Cousineau, 2018]. Finally, Cohen's d_z was obtained by dividing the mean difference by the standard deviation of differences. The resulting effect sizes indicate whether AI condition had a positive, negative, or negligible impact on participants' ratings.

Cohen's d_z effect sizes follow conventional small, moderate, and large thresholds

The interpretation of Cohen's d_z follows conventional effect size guidelines, with small, moderate, and large thresholds. Since the response scale in this study ranged from -2 to 2, these thresholds remain applicable:

- $|d_z| < 0.2 \Rightarrow$ Very small effect (practically negligible)
- $|d_z| \approx 0.3 \Rightarrow$ Small effect (noticeable but minor)
- $|d_z| \approx 0.5 \Rightarrow$ Moderate effect (meaningful impact)
- $|d_z| > 0.8 \Rightarrow$ Large effect (strong influence of AI condition)

The sign of Cohen's d_z also provides valuable information for the results of the After-Task Questionnaire:

- **Positive** d_z: The AI condition resulted in higher ratings compared to the No AI condition.
- **Negative** d_z: The AI condition resulted in lower ratings compared to the No AI condition.
- **Near zero** d_z: There was no substantial difference between conditions.

Therefore, Cohen's d_z values in this study help determine not only the *strength* of AI's impact but also its *direction*, whether AI had a positive, negative, or negligible influence on participant responses [Goulet-Pelletier and Cousineau, 2018].

In the following subsections, the results for all four sections of the after-task-questionnaire will be presented.

Collaboration Dynamics Results

In the collaboration dynamics section of the after task questionnaire, the participants gave lower ratings for the AI condition. Open communication in the group received

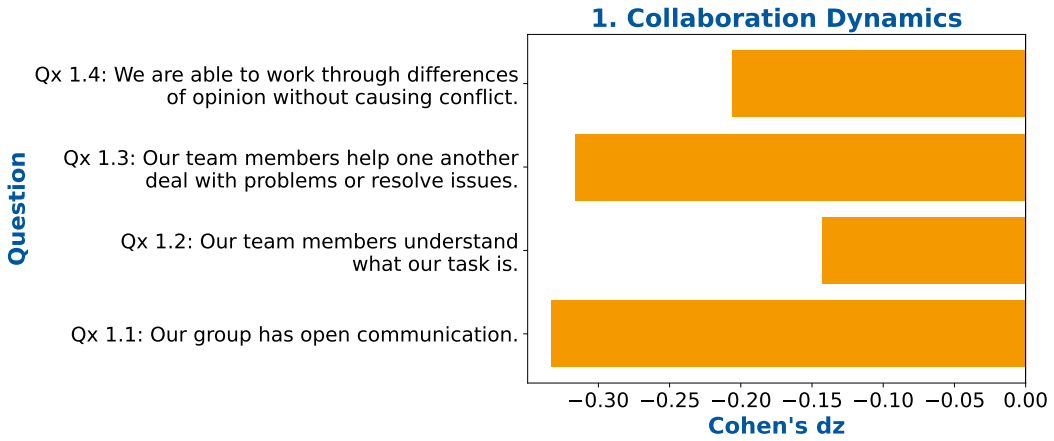


Figure 4.7: Cohen's d_z Effect Sizes for Collaboration Dynamics Section

Collaboration dynamics were rated lower in the AI condition, with small effect sizes for communication and problem-solving

lower scores in the AI condition (d_z = -0.33). Similarly, helping team members resolve problems was rated **lower** in the AI condition (d_z = -0.32). A slightly smaller effect was found for the statement "*We are able to work through differences of opinion without causing conflict.*", in which the participants gave the AI condition a **slightly lower** rating (d_z = -0.20). Meanwhile, there was not a large difference between the conditions regarding the group members understanding their task (d_z = -0.14). In Figure 4.7, the Cohen's d_z effect sizes for the collaboration dynamics section can be found.

Task-Specific Creativity Perception Results

Task-specific creativity was self-rated as more predictable in the AI condition, with small effect sizes for other creativity dimensions

In the task-specific creativity perception, the participants were asked to rate the creative output of their marketing tasks, which consisted of creating a slogan, a product name, drawing the product and logo, and also writing a few selling points. Their ratings were performed on a scale of -2 to 2, where the poles represent opposite adjectives. The creative output was rated as more predictable in the AI condition on the scale of predictable (-2) to unexpected (2), with d_z = -0.47. However, all remaining five adjective pairs had a negligible effect size. conventional (-2) vs. original (2) had d_z = -0.15, while impractical (-2) vs. practical (2) had d_z = 0.12. The remaining three pairs, unfocused vs.

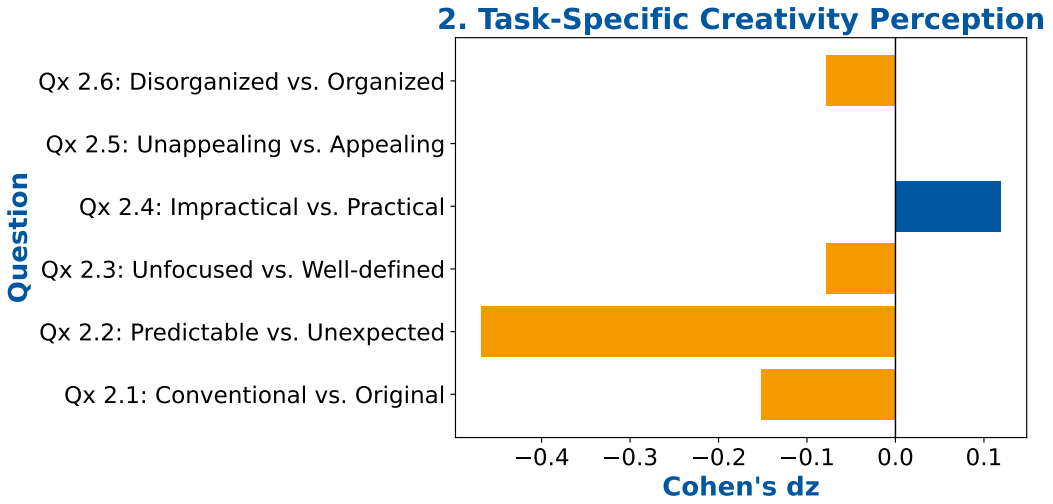


Figure 4.8: Cohen's d_z Effect Sizes for Task-Specific Creativity Perception Section

well-defined, disorganized vs. organized, and unappealing vs. appealing, had no meaningful difference in ratings. In Figure 4.8, the Cohen's d_z effect sizes per question for the task-specific creativity perception section can be found.

Emotional and Cognitive Load Results

This part of the questionnaire was adopted from the NASA Task Load Index (NASA-TLX). It is used for assessing the participants' perceived workload during a task. We used four out of the six original questions.

The self-reported mental effort to complete a task was **much lower** in the AI condition (d_z = -1.02). Additionally, the effort required to complete a task was **noticeably lower** in the AI condition (d_z = -0.55). However, the ratings for feeling **time pressure** and **satisfaction with performance** were identical in the AI and No AI conditions (d_z = 0). In Figure 4.9, the Cohen's d_z effect sizes for the emotional and cognitive load section can be seen. To ensure consistency in the interpretation of effect sizes across different sections, we inverted the values for the *emotional and cognitive load* section, specifically questions 3.1, 3.2 and 3.4. By inverting those values, a positive Cohen's d_z now consistently re-

AI reduced mental effort and task completion effort, with no effect on time pressure or satisfaction

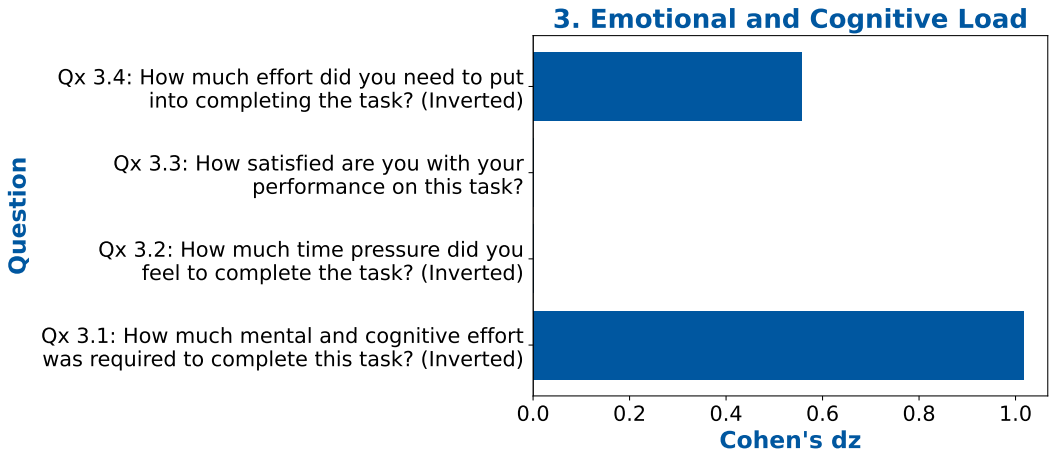


Figure 4.9: Cohen's d_z Effect Sizes for Emotional and Cognitive Load Section

flects a beneficial effect (AI reducing effort and cognitive load), while a negative Cohen's d_z suggests a unfavorable effect (AI raising effort and cognitive load). This has made comparison easier with the rest of the parts where positive values suggest an improved experience.

AI reduced cognitive workload, as shown by lower NASA-TLX scores in the AI condition

Furthermore, the average workload (NASA-TLX) was M = 44.10 (SD = 11.87) for the No AI condition and M = 35.42 (SD = 11.02) for the AI condition. Scores on the NASA-TLX usually range from 0 to 100. The larger the value, the greater the cognitive workload. In this study, participants first rated workload from 1 to 5 and then converted to the standard 0 to 100 range for comparison. The lower scores on the NASA-TLX in the AI condition indicate that participants experienced less workload when they used AI. However, one can see that individual experiences varied by looking at the standard deviations. The averaged NASA-TLX scores for both conditions can be seen in Figure 4.10.

Digital Tool's Role and Perceived Usefulness Results

This section of the questionnaire has the same 7 questions for both conditions, with 3 additional questions for the AI condition.

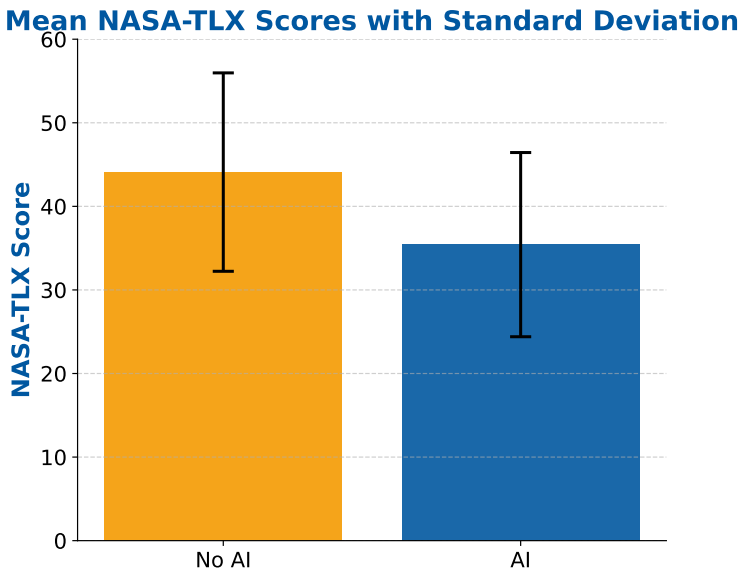


Figure 4.10: NASA Task Load Index Scores

Firstly, team collaboration was rated **lower** for the AI condition (d_z = -0.47), as well as the appreciation of the creative input of one's team members (d_z = -0.24). Moreover, the statement "*I could think independently.*" was rated *lower* for the AI condition, with d_z = -0.40. Teams also reported working *slower* with AI (d_z = 0.30). Participants reported a **slightly higher** enjoyment of their task (d_z = 0.14) during the AI condition. Regarding their need for more help during the task and their level of inspiration, the **difference was negligible**. In Figure 4.11, the Cohen's d_z effect sizes per question for the digital tool's role and perceived usefulness section can be found.

AI condition led to lower collaboration and independent thinking but slightly higher task enjoyment

In addition, the Cohen's d_z value for each question can be seen in Figure 4.12.

Regarding the remaining questions, which were only asked for the AI condition, participants reported a mean score of -0.33 (SD = 0.94, min = -2, max = 1) for feeling distracted by the digital tool. The negative mean suggests that most participants did not strongly feel distracted by AI.

Participants did not necessarily feel distracted by AI

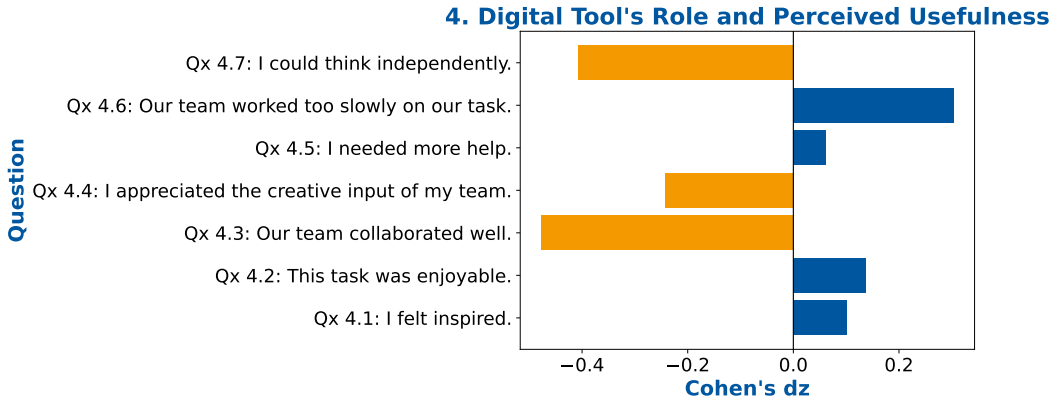


Figure 4.11: Cohen's d_z Effect Sizes for Digital Tool's Role and Perceived Usefulness Section

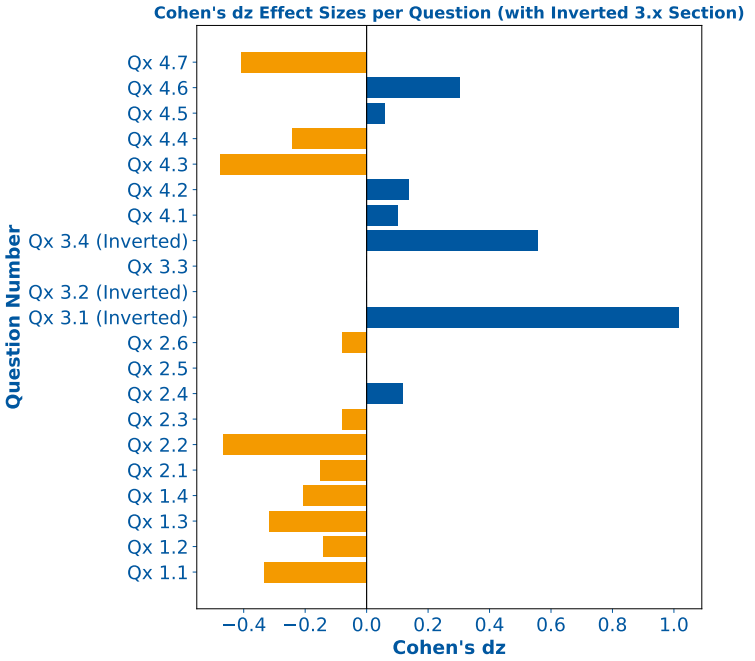


Figure 4.12: Cohen's d_z Effect Sizes per Question

Participants felt confident in using AI, and did not report it as limiting creativity on average

The participants rated their confidence in knowing when to use the digital tool with mean score of 0.83 (SD = 0.83, min = -2, max = 2). The positive mean indicates that participants generally felt confident in deciding when to use AI. For the question asking if the digital tool limited creativity,

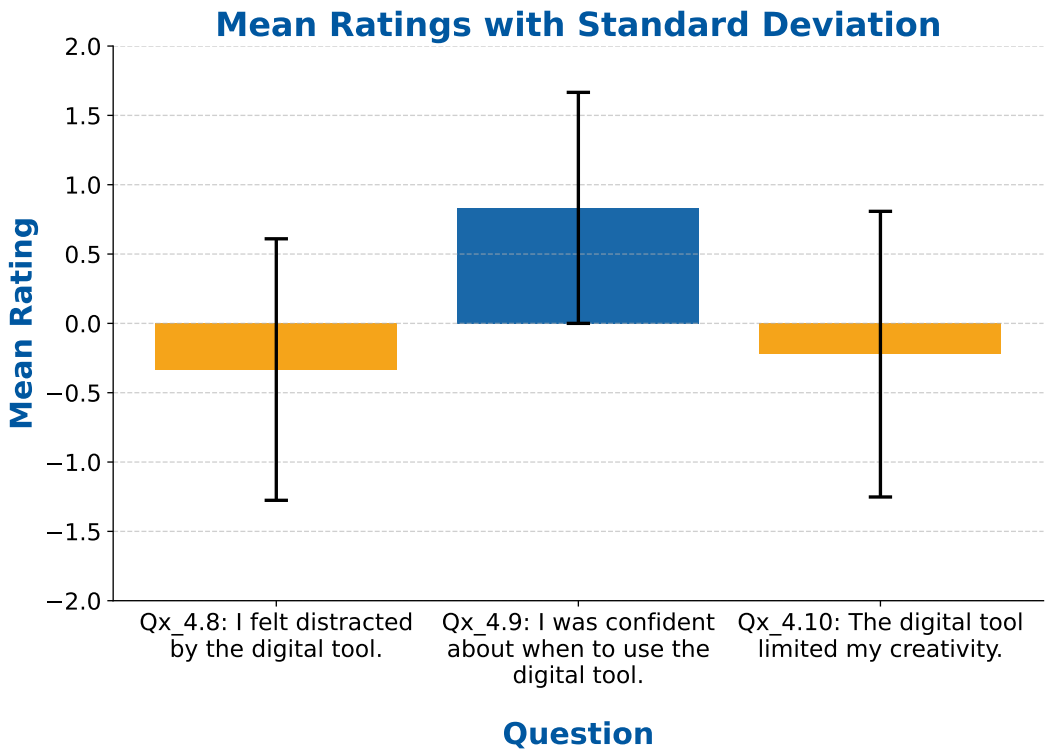


Figure 4.13: Mean Ratings of Qx 4.8 to 4.10 with SD

the mean score was M = -0.22 (SD = 1.03, min = -2, max = 2). The negative mean suggests that on average, participants did not perceive AI as significantly limiting their creativity, though the standard deviation (SD) indicates some disagreement among participants. The mean values along with the standard deviation for questions Qx4.8/9/10 can be seen in Figure 4.13.

4.5.4 Final Questionnaire

Product preference was nearly split, with a slight favoring of the Smart Plant Pot

The Final Questionnaire consists of 6 questions. In question **F1**, when asked which product they preferred marketing, 9 participants chose the Smart Plant Pot, while 8 preferred the Glow-in-the-dark Umbrella. Only one participant showed no preference, suggesting that opinions were nearly evenly split between the two products.

Among the participants who chose the **Glow-in-the-dark Umbrella**, common reasons were that it was more fun product for brainstorming, more creative, and general familiarity with the product. Some also included that they would use the product themselves, and preferred that outcome of the marketing task. For those who preferred the *Smart Plant Pot*, reasons centered on the product being more exciting and offering more creative possibilities. Some participants said it seemed more functional and practical, while others mentioned liking its versatility and the marketing potential it had. In question **F2**, when asked which condition they preferred for completing the tasks, 7 participants favored using AI, while 4 preferred working without AI. Meanwhile, 7 participants expressed no preference, indicating that opinions on AI assistance were somewhat divided. Those who preferred working **without AI** said brainstorming was more fun, allowing them to think outside the box and feel more personally involved in the creative process. Some also said the AI condition reduced creativity. On the other hand, participants who favored the **AI condition** found it easier to complete the task, with some mentioning that AI helped with inspiration and made generating product names and slogans more efficient. A few participants highlighted that AI allowed them to think beyond their usual ideas, though one noted their possible bias toward AI due to preferring the product. Meanwhile, those who had **no preference** acknowledged that both AI and No-AI conditions had advantages and disadvantages. Some noted that using AI provided more options but reduced creativity, whereas not using AI boosted creativity but reduced options.

Preferences for AI or No-AI conditions varied, with some valuing AI's efficiency and others enjoying the creative freedom without AI

The results for questions F1 and F2 can be seen in Figure 4.14.

In question **F3**, most participants (10) reported feeling that their group produced the most creative results without AI , while only 4 preferred the AI condition, and 4 saw no difference. Similarly in question **F4**, when it came to teamwork, the majority (9) felt both conditions were similar, while 7 believed their group worked best without AI, and just 2 preferred the AI condition.

Most participants felt their group was more creative and worked better without AI

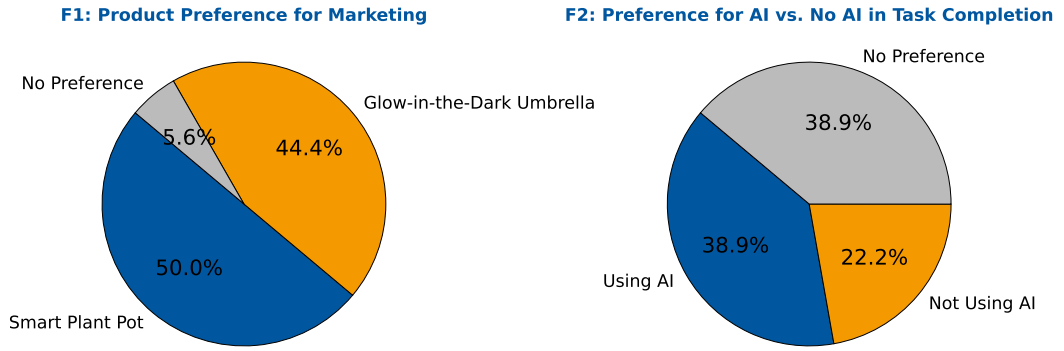


Figure 4.14: Product Preference (left) and Condition Preference (right)

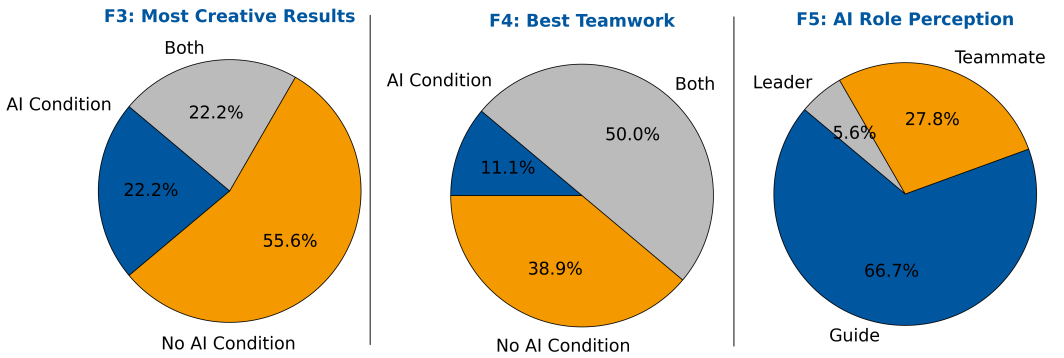


Figure 4.15: Most Creative Results (left), Best Teamwork (middle) and AI Role Perception (right)

Also, when asked about the AI's role during the task in **F5**, most (12) participants perceived it as a guide, while 5 saw it as a teammate, and only 1 participant considered it a leader.

Most participants viewed AI as a guide, with few perceiving it as a teammate or leader

The results for questions F3, F4 and F5 can be seen in Figure 4.15.

The 6th and final question of all questionnaires, **F6**, was an optional question, in which the participants were asked "*How would you briefly describe your overall experience working on the tasks?*". Many participants chose to answer this question, with only four opting out. A common expression among the participants was that the marketing experience was fun and creative, and some also added that it

was *"good to test the creativity"*, *"fun to see how creativity works"*, and that they had their *"cognitive cogs running"*. Additionally, a few participants gave a detailed response. *"It was interesting to have a direct comparison of how different the workflow is in both tasks and the results that get produced from each one."* said one of the participants, while another contributed a positive and also negative impression of their experience: *"I had fun working in a group setting. The products were thought-provoking. The AI added an extra layer that both helped the group, but also took away from the truly unique ideas"*. One participant also said that the AI felt like a teammate, because it suggested things similar to how a group member would join in on the brainstorming. Similarly, another participant enjoyed getting input from their group members along with the AI suggestions. Finally, one participant mentioned that one product felt more difficult to market than the other, and that their answers could potentially biased due to that fact.

Participants found the tasks fun and creative, with mixed opinions on AI's impact on uniqueness and group dynamics

Chapter 5

Study 2: Online Survey

It was decided during the study design process, to also conduct a second smaller survey-based study, in which the creative output of the first study could be rated and ranked externally by people who did not take part in Study 1. The motivation behind this study was to compare self-perceived rating of the creative output with the external ratings in both the AI and No AI conditions. Thereby answering one of the research questions:

Study 2 was conducted to compare self-perceived and external ratings of creativity

RQ1.4: How does AI impact how others perceive a group's creative result?

5.1 Preparation

Study 2 was conducted as a survey, which was created through the platform **SoSci Survey**[1]. It consisted of an informed consent form, an introduction explaining the survey, and a section explaining the nature of the marketing task from Study 1. There was no mention of the AI and No AI conditions in the survey. The survey was constructed into four sections. In the first section, the participants had to rate each group's entire creative result for the **Smart Plant Pot** product. For this section, the same questions from

Study 2 explained

[1] `https://www.soscisurvey.de/`

the Task-Specific Creativity Perception section from Study 1 were used, which can be seen in Figure 4.3 from the previous chapter. Afterwards, the participants were asked to rank the creative results from best (1) to worst (6). The were also asked to justify their ranking briefly. For the second product, we created the third and fourth sections as exact duplicates of the first and second sections, respectively, but for the **Glow-in-the-dark Umbrella** product. In every one of the four sections, the listed creative outputs were listed in random order to reduce potential order bias. The survey can be seen in Appendix A.

5.2 Results: Study 2

External raters gave higher creativity ratings to AI-assisted ideas, particularly for clarity and structure

The survey was completed by 20 people in the span of a week. The results show higher ratings for the AI-assisted marketing results, despite the participants of Study 2 not knowing about the different conditions in Study 1. This can be seen in Table 5.1. The table displays the mean creativity ratings for both products combined. Ratings were given on a **scale from -2 to 2**, with higher values indicating positive evaluations. Overall, ideas generated with AI assistance received higher ratings across all dimensions, when compared to the ideas generated without AI. The largest difference was found on the Unfocused vs. Well-defined scale, where AI-assisted ideas were rated as substantially **more well-defined** (M = 0.683) than ideas generated without AI (M = 0.35). This suggests that participants found AI-supported ideas clearer and more structured. Altogether, these results indicate that across both products, AI assisted creative results had **higher ratings** regarding their **definition, originality, practicality, appeal** and **organization**.

Condition	Originality	Unpredictability	Definition	Practicality	Appeal	Organization
AI	0.77	0.46	0.69	0.93	0.70	0.84
No AI	0.63	0.44	0.35	0.71	0.52	0.54

Table 5.1: Overall Creativity Ratings for Both Products (AI vs. No AI)

Meanwhile, Table 5.2 demonstrates the creativity ratings for both Smart Plant Pot (P), and the Glow-in-the-dark Umbrella (U) separately, while also comparing the AI condition with the No AI condition. Across *both products*, **AI-supported** ideas received higher ratings across most dimensions, particularly for **originality, definition, and organization**. For the *Smart Plant Pot*, AI-assisted ideas were rated as more **original, practical, appealing, organized** and **better defined** than those created without AI. For the *Glow-in-the-Dark Umbrella*, AI-supported ideas were seen as more **original** and **unpredictable**. However, the umbrella products from the No AI condition were rated as **more appealing** and **more organized**. Regarding definition and practicality, the umbrellas were rated very similarly for both conditions. Overall, AI assistance positively impacted both products, though to varying degrees.

AI-assisted ideas scored higher in originality, definition, and organization, with varying impacts across the two products

5.2.1 Ranking

The survey also resulted in the ranking of the products from best to worst. For the Glow-in-the-dark Umbrella, AI-assisted outputs ranked in both the 1st and 3rd positions, while an non-AI version was placed 2nd. For clarity and relevance, only the top three rankings are reported here.

The top-ranked Glow-in-the-dark Umbrella design was AI-assisted, but rankings varied across conditions

Participants justified ranking the Glow-in-the-Dark Umbrella concepts based on creativity, practicality, and uniqueness. Simpler designs were preferred, as "*less complicated is better,*" and customizable glowing features stood out as "*something unique.*" The eco-friendliness and safety features were also important factors, with one participant highlighting "*the storm-proof aspect of Lumibrella*" and the appeal of "*customizable light patterns, solar energy and rain activation*" in Starella. Target group relevance also played a role, with one person noting that Funbrella is "*a very nice and considerate idea for kids*". The creative outputs of each group can be found in Appendix B.

Rankings favored simple, unique, and eco-friendly designs with strong target group relevance

Meanwhile, for the Smart Plant Pot, AI-assisted outputs received both the 1st and 2nd ranks, while the 3rd place was

not AI-assisted. This suggests that participants generally preferred the AI-supported designs for the plant pot task.

AI-assisted Smart Plant Pot designs ranked higher, valued for innovation, clarity, and practicality

For this product, participants reported ranking based on creativity, practicality, and clarity of features, with several highlighting the appeal of BotaniQ for its "*combination of innovation, practicality and clarity through solar energy, self-watering and app control*". Memorable names and catchy slogans also influenced decisions, though some were seen as "*cringe*". Designs that were well-explained, visually clear, and easy to understand were rated higher, while some concepts were criticized for being unclear or offering fewer features. Overall, participants favored products that balanced originality with usefulness and ease of use.

Item	Originality	Unpredictability	Definition	Practicality	Appeal	Organization
P (AI)	0.88	0.35	1.08	1.20	0.95	1.18
P (No AI)	0.75	0.56	0.29	0.65	0.41	0.39
U (AI)	0.71	0.51	0.49	0.80	0.56	0.68
U (No AI)	0.38	0.20	0.48	0.82	0.73	0.85

Table 5.2: Creativity Ratings for Plant Pot and Umbrella (AI vs. No AI)

5.3 Comparison of Self and External Ratings Across Creativity Dimensions

Comparison of self-ratings and external ratings highlights potential biases in creativity assessment

Table 5.3 compares the mean self-ratings given by the groups who participated in Study 1 to the external (Ext.) ratings given by others in Study 2. Ratings were collected across six creativity dimensions: Originality, Unpredictability, Definition, Practicality, Appeal, and Organization. The self-ratings represent how participants evaluated their own ideas, while the external ratings reflect how other participants evaluated those same ideas. This comparison helps identify potential biases in self-assessment and how creators perceive their work compared to external evaluators, in both the AI and No AI conditions. This is also relevant for answering the following research questions:

RQ1.3: How does AI impact how group members perceive their group's creative result?

RQ1.4: How does AI impact how others perceive a group's creative result?

Interestingly, the pattern of ratings is similar across both the AI and NO AI conditions. Participants on average rated their own ideas higher in practicality, organization, definition and appeal compared to external evaluators. Meanwhile, external ratings were higher for originality and unpredictability.

Self-ratings favored practicality and organization, while external raters valued originality and unpredictability

Condition	Originality	Unpredictability	Definition	Practicality	Appeal	Organization
AI (Self)	0.11	-0.56	1.06	1.28	1.06	1.00
AI (Ext.)	0.77	0.46	0.68	0.93	0.70	0.84
No AI (Self)	0.39	-0.06	1.11	1.17	1.06	1.06
No AI (Ext.)	0.62	0.44	0.35	0.71	0.52	0.54

Table 5.3: Comparison of Self and External Ratings Across Creativity Dimensions

Chapter 6

Discussion

In the upcoming chapter, we discuss our findings from Study 1 and Study 2, and we provide a qualitative and quantitative analysis of the results.

6.1 Discussion of Qualitative Results

During Study 1, we had co-located groups work together on creative tasks. For this, we collected data through video recordings of each session, as well as screen recordings. We also collected a multitude of quantitative and qualitative data through the questionnaires.

Study 1 qualitative data: video, screen recordings, and questionnaire responses

We decided to analyze our qualitative data through Reflexive Thematic Analysis (RTA). Reflexive Thematic Analysis is a theoretically adaptable and flexible approach that is widely used in psychology [Braun and Clarke, 2006, 2019]. RTA also emphasizes the researcher's active role in constructing themes, as compared to other methods that require discovering pre-existing patterns within the data [Braun and Clarke, 2019]. This is a suitable approach for our study, since the analysis was conducted by a single researcher, whose interpretations and theoretical framework shaped the final themes.

Reflexive Thematic Analysis was used to construct themes from qualitative data

RTA involves coding data and constructing themes through researcher interpretation

The Reflexive Thematic Analysis method involves immersing oneself in the data and developing codes to capture interesting features, then constructing themes that highlight patterns that have a shared meaning across the dataset. Importantly, these themes are created through the researcher's active interpretation and reflexive engagement with the data [Braun and Clarke, 2019].

RTA is a flexible method for identifying meaningful patterns in qualitative data

According to Braun and Clarke [2006], thematic analysis provides an achievable entry point for researchers who are new to qualitative research, that also facilitates a rich and reflexive qualitative analysis. Reflexive Thematic Analysis involves the identification, analysis, and construction of patterns with similar meaning in qualitative data, such as textual feedback. The flexibility of the method allows researchers to tailor their analysis to their specific research questions, theoretical orientations, and emerging understandings of the data. While Reflexive Thematic Analysis is a flexible and recursive process rather than a strict procedure, it is typically guided by six broad phases that take the researcher from familiarization with the data to the final reporting [Braun and Clarke, 2006, 2019]. The phases are not in strict linear sequence, and it is possible to move back and forth between them throughout the analysis process.

RTA follows flexible phases, from data familiarization to interpretive reporting

1. **Familiarizing Yourself with Your Data**: During the first phase, we immerse ourselves with the data. This is achieved by reading or watching it several times. It helps us gain an intuitive and interpretive sense of the depth and breadth of the dataset [Braun and Clarke, 2006, 2019].
2. **Generating Initial Codes**: We conduct a dynamic coding process by creating codes that capture meaningful features of the data for our analysis. In RTA, coding is interpretive and reflexive, rather than a mechanical sorting of data [Braun and Clarke, 2006, 2019].
3. **Constructing Themes**: Rather than simply grouping codes, we engage deeply with the data to construct themes that show patterns of shared meaning. This is guided by our evolving interpretations of the data [Braun and Clarke, 2019].

4. **Refining and Developing Themes**: We continue our process by reviewing the themes iteratively. By doing this, we ensure that the themes form coherent descriptions and narratives about the data. We achieve this by critiquing the themes to explore their depth and their analytical significance, instead of looking for strict coherence [Braun and Clarke, 2019].
5. **Defining and Naming Themes**: We elaborate the themes further, and attention is given to their fundamental concept that gives meaning to each theme. The goal is to produce themes that tell a deep analytical story rather than simply summarizing data [Braun and Clarke, 2019].
6. **Producing the Report**: In the final phase, we write a rich and interpretive analysis that presents a strong narrative supported by data extracts, for example as quotes. The analysis highlights the researcher's interpretative role in formulating the findings rather than treating themes as objective facts [Braun and Clarke, 2019].

We started our qualitative analysis journey by coding the segments of each video using the trial version of MaxQDA, which is a software that assists with qualitative and mixed-methods analysis.[1] Each video was initially cropped to only contain the segments in which the groups performed their tasks in the AI and No AI conditions. Each cropped video was between 50 and 70 minutes long. We afterwards used a local research transcriber called *AixWhisper* to create the transcripts for our cropped user study videos.

MaxQDA and AixWhisper were used for video coding and transcription

Initially, the videos were reviewed and a few initial codes were created. However, during the active coding process, many new codes were created and assigned to specific video segments. A few examples of codes created during this process include *'Rhyming for Slogan'*, *'Suggesting to Prompt AI'*, *'Telling Group Member What To Write on Worksheet'* and *'Googling Information'*. Each code was also assigned a 'memo', in which its meaning is briefly described. The coding process involved active interpretation, as codes

Coding process generated interpretive labels to capture group interactions and behaviors

[1] `https://www.maxqda.com/`

were developed to capture interesting features of the data that reflected our evolving understanding of the groups' interactions. The codes were not treated as objective facts, but rather as conceptual tools for capturing patterns of meaning within the dataset [Braun and Clarke, 2019].

Themes were constructed by interpreting relationships between codes and research questions

Once all of the codes had been created, we began the reflexive process of theme construction, where codes were examined for patterns of similar or shared meaning that could be brought together into coherent themes. This process involved actively interpreting how different codes related to each other and how they connected to the broader research questions. Themes were not simply summaries of content, but rather interpretive stories about the data that reflected both the participants' experiences and our theoretical and analytical lens [Braun and Clarke, 2019].

Codes were grouped into four clusters to explore patterns in AI, tasks, tools, and group interactions

During this analytic process, we grouped the codes into four clusters or themes: **AI-Related**, **Task-Related**, **Tool-Related**, and **Group Interaction and Discussion**. These clusters were working organizational groupings that allowed us to begin exploring patterns of meaning across the dataset. The clusters had varied sizes, with the smallest having 6 codes, and the largest cluster consisting of over 20 codes. These clusters were used as analytic tools to support our reflexive analysis of the data, rather than fixed components for the final structure of themes [Braun and Clarke, 2019].

Videos were coded by condition, with frequency data providing context for interpretive theme development

After all of the 6 of the videos were fully coded, they were split into two sections, AI Condition and No AI condition. We created a table summarizing how often each code appeared in each condition. While these frequencies provided some context, our analysis focused primarily on developing rich, interpretive themes that captured meaningful patterns across the data, rather than treating frequency as a measure of high importance [Braun and Clarke, 2019]. We also looked at the edge-cases, in which codes were not frequent, but quite relevant to the research questions. In addition, we examined and reviewed the themes in each group, in order to catch any interesting or significant patterns. We also revisited our coded videos, and the transcripts we had

generated for them, to find segments and quotes from our data that support our findings.

Throughout the analysis process, constant reflexive reflection was vital to our approach. This involved regularly questioning how our assumptions, expectations, and theoretical positioning shaped the development of codes and themes. In her paper, Validation of Qualitative Research in the "Real World", [Pyett, 2003] addresses the ongoing debate surrounding the concept of validity in qualitative research, arguing that validity should not be judged by traditional scientific criteria but by the transparency of the researcher's interpretive process. In line with this, our ongoing critical reflection ensured that the themes were not presented as objective truths, but rather as interpretive stories that we co-constructed through the interaction between ourselves and the data [Braun and Clarke, 2019]. In the upcoming sections we address the main research question and its four sub-questions.

Reflexive reflection ensured themes were transparent, interpretive, and not treated as objective truths

"RQ1: How do generative AI tools impact the creative process and outcomes in co-located group settings?"

6.1.1 AI-Related Theme

The **AI-Related** theme reveals how participants interacted with AI during the user study, which includes patterns of trust, skepticism, and selective engagement. Although groups often made use of AI as a tool for generating content, they also recognized its limitations. This resulted in a dynamic process of acceptance, modification, and occasional frustration.

Participants engaged with AI dynamically, balancing trust, skepticism, and adaptation

A striking pattern across the groups was that participants frequently relied on the AI to generate ideas. This is evident in the high frequency of "*Suggesting to Prompt AI*", which occurred multiple times per group. In addition, groups who had the AI condition as their second condition prompted AI almost directly into their session (10 to 20 seconds into the session). This suggests that participants viewed AI as the first step to idea generation, when avail-

Participants frequently relied on AI for idea generation, especially when uncertain

able. They often defaulted to it when uncertain or needing inspiration. In the next paragraph, we exhibit an excerpt from the user study, in which participants are referenced as (Group, Position), where X, Y, and Z denote the left, middle, and right positions, respectively. For example, (03, X) refers to the leftmost participant in Group 3.

> **(06,Y)**: "Maybe the slogan could be like, *Glow and Go*." *(Laughter among the group.)* **(06,Y)**: "Or we can ask ChatGPT for a catchy slogan."

Participants selectively adopted AI-generated content

However, the interactions were not purely dependent on AI. The frequent occurrence of "*Selective Adoption of AI*" highlights that participants did not accept AI-generated content uncritically. Instead, they chose specific outputs to refine, modify, or combine with their own ideas.

> *(The group is discussing AI-generated product names for their Smart Plant Pot, such as PotPal, SmartSprout, and Planta.)*
> **(01,Z)**: "How about PlantPal?"

AI was seen as a useful assistant, though its limitations were noted in most groups

This suggests that AI was often perceived as a collaborative assistant whose contributions required human judgment. In addition, participants portrayed more instances of "*Liking AI Output*" than "*Disliking AI Output*," nearly at a 2:1 ratio. This could potentially show that, in general, AI-generated content was useful, engaging, or at least satisfactory for most participants. At the same time, "*Perceiving Limitations of AI*" appeared in 5 out of 6 groups. This could indicate that while AI was appreciated, it was not always seen as producing high-quality or relevant suggestions. Participants sometimes found its outputs generic, inaccurate, or needing significant revision.

> *(The group is looking at an AI-generated image.)*
> **(04, Y)** *(sarcastically)*: "Well, that is definitely interesting."
> **(04, X)**: "I think we can just do it better."
> **(04, X and Z)**: "Yeah."

This suggests that while AI was viewed as an efficient brainstorming tool, its perceived quality was inconsistent, leading participants to evaluate outputs before adopting them.

AI was frequently used for image generation, valued for speed and variety

In addition, a particularly strong pattern emerged in the use of AI for image generation. Four out of six groups used AI to generate images for their visual task, with most generating at least five images per session. This suggests that participants found AI especially valuable for visual tasks. This is possibly due to the speed and ease of generating multiple versions.

Edge case: Information overload from excessive AI suggestions

Moreover, two edge cases emerged in the AI-Related theme. The first edge case is "*Information Overload (AI)*". In one case, a participant expressed being overwhelmed by the amount of AI-generated suggestions.

> *(The group is looking at an AI-generated text.)*
> **(03, Y)**: "It's so much that it's too hard to decide."

Edge case: Reflecting on reliance on AI

The other edge case observed in this theme is "*Reflection on AI Dependency*". Only one group explicitly reflected on their increasing reliance on AI, suggesting that most participants engaged with AI without questioning its impact on their creative autonomy.

> *(The group is writing the last words on their worksheet during the second condition and begins discussing their experience with the study.)*
> **(01, Y)**: "The moment you're given some kind of tools, your brain shuts down and says, 'I'm gonna use the tools.'
> But the moment you have no possibility to use the tools, it's like your potential comes out more."

Many groups used AI without reflecting on their dependency. Potential area for future work

These cases indicate that while AI was heavily integrated into the creative process, most groups did not step back to reflect on their level of dependence on AI during their task, highlighting a potential area for further study.

6.1.2 Task-Related Theme

The **Task-Related** theme inspects how participants engaged with their tasks across both AI and No AI conditions. Our data shows that some tasks were significantly affected by AI's presence, while others were generally unaffected.

AI reduced group discussions on product technicalities, shifting focus away from refining product details

One of the most noticeable difference was the decline in "*Group Discussion About Product Technicalities*" when AI was available. In the No AI condition, participants often debated and refined the specifics of their product, i.e. what it does, how it works, and why it would appeal to consumers. However, in the AI condition, such discussions occurred at only about ¼ of the frequency observed in No AI.

> *(Explaining to group members the concept of self-priming and having an irrigation system during No AI condition.)*
> **(05, Z)**: "It analyzes the moisture of the water, and based on that, it can actually kind of like predict how much water it needs."

AI reduced in-depth product discussions

This suggests that when AI was available, participants were less likely to discuss the product's features in depth. This could be due to relying on AI-generated ideas rather than refining their own ideas.

Idea generation was more frequent without AI, especially for naming and slogans

In addition, a large difference between conditions occurred in "*Proposing a New Idea*", which was nearly three times as frequent in the No AI condition compared to the AI condition. This suggests that without AI, participants engaged more actively in idea generation. They relied on their own creativity, rather than relying on AI for ideas. However, this was not the case for all sub-tasks. For naming and slogan tasks, participants in the No AI condition suggested ideas three times more often than in the AI condition. These tasks require creativity and linguistic play (e.g., rhyming, wordplay, cultural references), which participants may have found more engaging or intuitive when done manually.

> *(The group is trying to create a slogan in the No AI condition.)*
> **(01, Y)**: "Hello little fella, grow with the Funbrella."
> *(Demonstrating rhythm)*: "It goes up, down, up, down."

This pattern suggests that AI may help with idea generation but can also reduce spontaneous creativity in tasks that rely on linguistic play and conceptual flexibility. In addition, AI's impact on task engagement was task-dependent. It **significantly influenced** ideation-based activities like creating the product name and slogan, while having **minimal effect** on fact-based tasks such as creating selling points.

AI reduced spontaneous creativity in linguistic tasks but had minimal effect on fact-based tasks

Moreover, across both conditions "*Recalling Previous Decisions*" occurred at nearly identical rates, suggesting that AI did not fundamentally alter how participants revisited or reinforced earlier choices. The consistency in decision recall across conditions suggests that while AI influenced how ideas were generated and refined, it did not heavily disrupt the teams' working structure.

AI did not strongly impact how participants recalled and reinforced decisions

Another notable pattern emerged regarding when the groups first prompted AI. For all groups who had it as their second condition, they used it almost directly (10 to 20 seconds into the session). However, some groups who had it as their first task waited a few minutes to first prompt AI.

Groups prompted AI almost directly when it was their second condition

6.1.3 Group Interaction and Discussion Theme

The **Group Interaction and Discussion** theme directly addresses the research question **"RQ1.1: How do AI tools influence collaboration dynamics in the creative process of groups?"**. We specifically explore how the availability and use of AI impacts interactions in the groups, role distributions, and decision-making processes within collaborative creative tasks.

This theme explores AI's impact on collaboration, roles, and decision-making in groups

No AI condition had more disagreements and clarifications about task details

Participants in the No AI condition demonstrated more frequent cases of "*Disagreement About Task Details*" and "*Concept Clarification to Other Group Members*", in which each occurred approximately twice as often compared to the AI condition.

> *(No AI condition, thinking of product name for the glow-in-the-dark umbrella.)*
> **(01, X)**: "Maybe like, LumBrella."
> **(01, Y)**: "I don't think anyone would... If you say something like GlowBrella, people would get what you mean with it. But LumBrella, people would be like, 'What are you talking about?'"

No AI groups had more direct instructions and silent reflection, indicating deeper engagement

Additionally, the behavior of "*Telling Group Member What To Write on Worksheet*" was observed three times more often when there was no AI support. Similarly, periods of "Silent Thinking" were three times more common in the No AI condition. These patterns suggest that, without AI assistance, group interactions involved deeper interpersonal engagement, including negotiation, clarification, and reflection.

> *(Participant Z is writing the selling points during the No AI condition.)*
> **(06, Y)**: "You can write down 'manageable via app'."

AI condition had more multitasking and occasional role-switching within groups

On the contrary, the presence of AI appeared to bring our different collaborative behaviors. Participants showed more instances of "*Within Group Multitasking*", in which group members worked simultaneously on separate subtasks. This was observed approximately twice as often in the AI condition, compared to the No AI condition. Additionally, group members only switched roles during the AI condition. For example, two participants switching between roles of typing on the keyboard and writing on the worksheet. However, this only occurred during two of the six sessions.

Additionally, the frequency of "*Asking Teammates for their Opinion*" remained consistent across both conditions. This shows that this behavior was a stable component of group collaboration, and it was not significantly affected by the availability of AI.

Asking for teammates' opinions remained stable across both conditions

> *(A group is discussing potential colors for the umbrella during the No AI condition.)*
> **(04, X)**: "What is it in your mind? Which color?"

Overall, these findings suggest that AI availability significantly influences collaborative strategies. It reduces explicit negotiation and introspective participation while encouraging dynamic multitasking. However, certain interpersonal interactions, such as seeking teammates' opinions, remain stable across conditions, indicating fundamental collaborative behaviors persist regardless of AI presence.

AI reduces group negotiation and introspection, but core collaborative behaviors remain similar

6.1.4 Tool-Related Theme

The **Tool-Related** theme captures how participants chose and engaged with different tools, especially comparing the use of AI to traditional tools such as Google searches.

Participants interacted with AI nearly twice as frequently as they conducted traditional internet searches (e.g., Google). This could indicate that when AI was available, it became a primary tool for information gathering and ideation. Although Google was not used as often, it was used at least once in all of the No AI sessions. Furthermore, generating images with AI was strongly favored for the drawing task, occurring roughly twice as frequently as traditional image searches. Participants consistently generated multiple AI-produced visuals until they were satisfied with the output. This could be explained by AI's perceived convenience and rapidness in visual ideation compared to regular web searches.

When available, AI was the primary tool for ideation

These findings could indicate that participants viewed AI as more efficient or appealing compared to traditional

AI's efficiency influenced participants' workflow and decision-making

search methods, which significantly shaped their workflow and decision-making processes.

6.1.5 Group-Specific Discussion

Group dynamics were analyzed alongside group familiarity ratings from the background questionnaire

While constructing themes for the Reflexive Thematic Analysis process, we also noticed interesting occurrences in some of the groups. In attempt to explain the observed phenomena, we further investigated these specific group dynamics by also checking the results of the background questionnaire, in which the group members rated how well they knew their group members.

Excessive AI reliance in Group 2 led to concerns about limited idea exploration

Group 2 uniquely exhibited the behavior "*Suggesting not to Rely on AI*". One member explicitly noted a sense of excessive reliance on AI in their group, indicating teammates did not explore ideas outside the AI's suggestions. The same group member wrote the following in their After-Task questionnaire:

> **(02, Z)**: I felt that my teammates were relying too much on GPT. They only chose between the given options and did not consider ideas it did not generate, even when asking them specifically for other ideas outside ChatGPT.

This observation suggests that excessive reliance on AI may affect group satisfaction and points to the importance of balanced AI integration. Moreover, the over-reliance can be interpreted as social loafing, meaning that users who viewed AI as a partner often relied too much on its input [George, 1992; Luan, 2024].

Group 5's high group familiarity correlated with more disagreements and direct task management

Meanwhile, Group 5 exhibited notably frequent occurrences of interactions such as "*Defending Idea*", "*Telling Group Member What to Write on Worksheet*" and "*Stressing About Time*", especially prominent in the No AI condition. Group 5 also had the highest mean result among all groups for the question "*How well do you know your group members?*" in the background questionnaire. This

group's deeper interpersonal engagement could be linked to their strong familiarity with each other, potentially leading to more openly expressed disagreements and defenses [Muskat et al., 2022].

Another notable observation was in Group 3, which reported the lowest average for group members' familiarity with each other on a scale of 1 to 5 (M = 1, SD = 0). Group 3 often had an imbalance of work due to ignoring or rejecting input from a specific group member, particularly in the No AI condition. This dynamic could be explained by the group's low familiarity, suggesting that lower familiarity among participants might lead to exclusionary behaviors and reduced collaborative cohesion [Muskat et al., 2022].

Group 3's low familiarity correlated to work imbalance and reduced collaboration

6.1.6 Discussion of Qualitative Results From Questionnaires

The participants were asked the same open-ended questions in both After-Task Questionnaires for the AI and No AI conditions. In the following sections we applied Reflexive Thematic Analysis to analyze the answers for the open-ended questions.

Open-ended questionnaire responses were also analyzed using RTA

"What were your overall thoughts or impressions about working in this condition?"

Among the themes observed in the responses from the **No AI condition** were "*Independence and Internal Creativity*", in which participants highlighted valuing independent thinking and generating ideas naturally without AI assistance. Some even expressed feeling pride in their creative output. Another theme observed was "*AI awareness and comparison*", especially among participants whose first task was AI-assisted. For example, someone expressed they were "*always thinking about how ChatGPT would be helpful for ideating*", while another participant expressed that "*using AI made (their) previous task more original*". In addition, multiple participants pointed out how the process was "*harder without AI*".

No AI condition responses emphasized independence, pride, and AI comparison

In contrast, the answers for the same question in the **AI condition** highlighted how the task was "*easier, less fun, (and) less creative*". Multiple responses further highlighted feeling dependency on AI, such as "*It feels like I was stripped away from my own originality and creativity due to comfort of having ChatGPT at disposal*". It was also a common theme for participants to feel less engaged in the task due to AI doing much of the work. Among the more positive responses in the AI condition is that AI was helpful for generating fresh perspectives and providing inspiration, especially for slogan and name creation tasks. Many additionally mentioned how the AI tool was helpful for managing and reducing the effort needed to complete the tasks.

AI condition responses noted ease but less creativity and engagement, with AI aiding inspiration and efficiency

Overall, while the No AI condition fostered greater independence and engagement despite being more challenging, the AI condition provided convenience and inspiration but often at the cost of creativity and active participation.

No AI fostered engagement and creativity, while AI offered convenience

"What aspects of this condition did you find most helpful, if any?"

In the **No AI condition**, participants highlighted the benefits of collaboration and creativity. They emphasized the role of teamwork and discussion in idea generation. Working with teammates and exchanging thoughts helped them refine concepts, with one participant mentioning the usefulness of having "*teammates to bounce off of as well as googling for a quick fact check-up.*" Some also appreciated the ability to "*streamline a single idea till it became better.*" Independent thinking and collaboration were recurring themes in the answers. Participants had to rely on their own cognitive efforts when AI was not at their disposal, which some found creatively rewarding, although it increased difficulty. As one participant said, "*it made us think for ourselves rather than rely on a finished result from an AI model*". Participants felt a sense of ownership over the creative process when AI was not present. Several participants also appreciated the absence of AI, describing the process as "*more creatively rewarding, albeit harder*" and emphasizing that there were "*less distractions*" compared to working with AI. Lastly, some

No AI condition encouraged teamwork, independent thinking, and a stronger sense of creative ownership

participants used external resources as a supplement in the absence of AI. Using Google searches or reference materials helped them compensate for AI. One participant mentioned that "*googling reference pictures made drawing easier.*" These responses suggest that while the absence of AI made tasks more challenging, it also encouraged stronger participation in teamwork and more independence.

In contrast, in the **AI condition**, participants mentioned efficiency and reduced effort as key benefits. They noted that tasks felt "*quicker, low effort*" and that AI provided "*speed and confirmation*" in idea generation. Many of them also appreciated the role ChatGPT plays in promoting creativity. One participant mentioned that "*ChatGPT gives many options for you to choose from*", while another found it helpful for "*generating slogans and coming up with new directions if stuck*". AI was also seen as a tool that allowed participants to evaluate options faster than they would on their own. Some emphasized the impact of AI on textual and visual creativity, for example, by using the image generator or brainstorming slogans. Although most found AI useful, a few responses suggested that teamwork was still vital for the collaborative process. One participant noted the combination of "*using ChatGPT and teamwork.*" Overall, the AI condition was appreciated for speed, efficiency, and idea generation, although it could also have reduced the cognitive effort typically needed for ideation.

AI condition was valued for speed, efficiency, and idea generation, though teamwork remained important

"What challenges did you face while working under this condition, if any?"

In the **No AI condition**, some participants struggled with their task, especially due to "*lack of knowledge about certain aspects of the product, e.g. what is neon vs. fluorescent*". Some also mentioned feeling a strong sense of time-pressure. Other participants also mentioned that differences in opinions and the time limit affected their creativity, making it harder to brainstorm freely. One participant noted: "*Difference of opinions, time-limit, not much creativity.*" Participants also often noted struggling with solving the specific tasks in marketing, such as creating a slogan, "*looking for sell-*

No AI condition posed challenges in knowledge gaps, time pressure, and marketing-specific tasks

ing points" and "*phrasing thoughts in a marketable fashion*". Meanwhile, others noted that without AI, they had to find alternative sources of inspiration. One expressed that "*failing to come up with ideas by ourselves made it necessary to think of alternative ways to find inspiration*".

AI condition led to over-reliance, choice overload, and challenges in prompting effectively

However, in the **AI condition**, participants expressed over-reliance on ChatGPT, for example "*I felt that my teammates were relying too much on GPT*". Others noted that AI's excessive options were overwhelming, making decision-making harder, for example: "*Too many options to choose from*". In addition, some participants seemed unsatisfied with the AI outputs at times. For example, one noted that "*sometimes the results were just really not what we were looking for, and looking for the right prompts slowed us down*". Multiple participants also noted finding prompting AI challenging, especially during image generation. For example: "*Logo generation in ChatGPT didn't work quite the way we wanted*".

To summarize the results of this question, on the one hand, in the No AI condition, participants struggled with knowledge gaps, time constraints, and independent idea generation. On the other hand, in the AI condition, they faced challenges with loss of creativity, overwhelming AI options, and difficulties in refining prompts and generating visuals.

"In what ways did you use the digital tool, and for what purpose?"

The final open-ended question was only asked during the AI condition, and the answers reflected many different perspectives. A common theme mentioned among the participants was using AI for ideating and brainstorming. One participant mentioned using AI "*almost every step of the way.*" Another similar theme was using AI for refining and improving the group's original ideas. The main two themes among the groups are **widespread AI usage** and **selective use of AI**. Most participants reported to have used selectively, by either not using it for each task, or by refining original ideas with AI. The remaining participants indi-

AI was used for brainstorming, refining ideas, or extensively throughout tasks

cated widespread AI usage throughout their tasks. For example, one participant said that they used AI *"for pretty much everything"*.

6.2 Discussion of Quantitative Results

For our quantitative data analysis, we followed the ethical data analysis guidelines of Yu [2020]. By following the guidelines, we objectively examined the quantitative data so that personal bias did not influence the interpretation. We also reporting the results without making changes to the results to fit expected or desired outcomes. To protect the anonymity of the participants, we removed any information that could identify them during our analysis and reporting. We also ensured that our findings did not go beyond the study context and scope.

Quantitative analysis followed ethical guidelines to ensure objectivity and participant anonymity

6.2.1 Discussion of Quantitative Results from Study 1

In Section 4.5.3, we calculated the results for the After Task Questionnaire, both for the AI and No AI conditions. This was achieved by calculating Cohen's d_z, rather than Cohen's d, due to the nature of the study design, in which we implemented within-subject design. Therefore, Cohen's d_z was calculated for each question to find differences between the two conditions.

Cohen's d_z was used to analyze After Task Questionnaire results in the within-subject design

Firstly, in the *Collaboration Dynamics* section, the AI condition showed lower scores for open communication and mutual assistance ($d_z = -0.33$), and slightly lower scores for working through differences without causing conflict ($d_z = -0.20$). This suggests that the teams perceived less active group interaction when AI was available.

AI condition showed lower scores for communication and mutual assistance, suggesting reduced perceived group interaction

In the *Task-Specific Creativity Perception* section, participants had to rate their creative output in both the AI and No AI conditions. In the AI condition, participants rated their creative output as more predictable ($d_z = -0.47$) and more

AI reduced perceived novelty and unpredictability in creative output

conventional ($d_z = -0.15$). The only attribute that was more positively rated in the AI condition was practicality ($d_z = 0.12$). Other attributes, such as organization, appeal, and definition, had very small effect sizes. This could imply that although AI can help refine ideas and make them more practical, it could also reduce the self-perceived novelty and unpredictability of creative work. The aforementioned findings address the second research question, **RQ1.3: How does AI impact how group members perceive their group's creative result?**

AI reduced cognitive effort but did not impact time pressure or confidence in output

Moreover, in the *Emotional and Cognitive Load* section, we investigated the emotional and cognitive load on the participants in each condition. Results show that AI assistance significantly lowered cognitive effort ($d_z = -1.02$) and overall effort ($d_z = -0.55$). The were no differences regarding time pressure or satisfaction with performance. The general NASA-TLX scores also dropped from (44.10) in the No AI condition to (35.42) in the AI condition. These results indicate that AI assistance can reduce mental and cognitive demand, making tasks feel less taxing for participants. However, since time pressure and satisfaction with performance remained unaffected, we can deduce that while AI reduces workload, it does not necessarily improve participants' perceived efficiency or confidence in their output.

AI condition lowered ratings for team collaboration, independent thinking, and appreciation of teammates' input

In the final section of the After-Task Questionnaires, the *Digital Tool's Role and Perceived Usefulness* section, we examined how AI influenced team collaboration, independent thinking, and overall task experience. Participants rated team collaboration lower in the AI condition ($d_z = -0.47$), along with their appreciation of team members' creative input ($d_z = -0.24$). The ability to think independently also received lower ratings with AI ($d_z = -0.40$). These findings confirm our qualitative findings that were presented in the previous section.

AI slightly slowed task completion but had a small positive effect on enjoyment

While teams reported working slightly slower with AI ($d_z = 0.30$), task enjoyment showed a small positive effect ($d_z = 0.14$). AI had no significant impact on the need for additional help or inspiration.

Regarding the questions that were only asked in the AI condition, participants generally did not feel strongly distracted by AI (mean = -0.33, SD = 0.94), and they felt confident in knowing when to use AI (mean = 0.83, SD = 0.83). When asked whether AI limited creativity, responses were slightly negative (mean = -0.22, SD = 1.03), indicating no strong perception of AI as a creativity constraint, though responses varied widely.

Participants felt confident using AI, with mixed perceptions on its impact on creativity

These results suggest that AI assistance had a mixed impact on collaboration and independent thinking. Participants felt that AI reduced team collaboration, their appreciation of creative input from others, and their ability to think independently. Although it did not significantly impact inspiration or the need for additional help. While AI slightly slowed down task completion, it also made the task slightly more enjoyable. Additionally, most participants did not feel strongly distracted by AI and were generally confident in knowing when to use it. Some participants perceived AI as limiting creativity, but the overall effect was minor. These findings indicate that while AI tools can support task performance, they may also change team dynamics and cognitive involvement.

AI altered team dynamics, reducing collaboration and independence while enhancing enjoyment

In Section 4.5.4, the results from the Final Questionnaire were presented. Regarding condition preference, most participants voted either in favor of **AI** or had no preference, with the minority preferring working on the task without AI. This shows that opinions were split; some participants liked AI for ease and inspiration, while others preferred the authentic creative experience of no AI. However, most participants voted in favor of the **No AI condition** regarding the creativity of their output. This indicates that many participants felt traditional brainstorming produced more original ideas. In addition, when asked which condition led to the best teamwork, most participants had no preference, with the second highest vote being in favor of the **No AI condition**. This suggests that most did not perceive AI to enhance teamwork, and some even believed it reduced group collaboration. Moreover, when asked about their perception of AI's role, most participants perceived it as a **guide**, a few saw it as a teammate, and only one person saw it as a leader. This indicates how AI was mainly seen as

Participants saw AI as a guide, preferring AI for ease. No AI was preferred for creativity and teamwork

a guiding tool rather than a leading force. This also directly answers the research question **RQ1.2:** How do participants perceive the role of AI in creative group work?

Participants found the task engaging, with AI seen as both helpful and limiting creativity

Furthermore, in the final question in this questionnaire, in which participants had the option to briefly describe their general experience, results show how participants found the marketing task engaging and creatively stimulating. It was described as "*fun to see how creativity works.*" Interestingly, AI was perceived as both helpful and limiting. One participant noted that it "*added an extra layer that both helped the group but also took away from the truly unique ideas.*" Some also saw AI as a teammate, suggesting it played an active role in brainstorming.

6.2.2 Discussion of Results from Study 2

In this section, we aim to address the research questions:

RQ1.3: How does AI impact how group members perceive their group's creative result?

RQ1.4: How does AI impact how others perceive a group's creative result?

Study 2 gathered external ratings on the creative outputs from Study 1

After conducting Study 1, we conducted a second study (Study 2) through a survey, in order to get external opinions on the creative outputs of the participants from Study 1. The survey consisted of rating and ranking each group's marketing results for the two products, the Glow-in-the-Dark Umbrella and the Smart Plant Pot.

Study 2: AI-assisted outputs were rated higher, especially for organization, definition, and practicality

The results, found in Section 5.2, reflect higher ratings for the AI-assisted products. When taking both marketed products into consideration, all six of the attributes were rated higher in the AI condition, which were originality, unpredictability, definition, practicality, appeal and organization. This can be seen in Table 5.1. AI-assisted creative outputs were especially rated as much more organized, well-defined and practical, when compared to the creative outputs from the No AI condition. A possible ex-

planation for why the AI-assisted results appeared clearer and more structured to external observers could be due to the systematic way AI tools generate ideas. They often provide fully formed phrases, bullet points, or lists that produce clear and organized ideas.

Study 2 suggests AI outputs may be rated lower when evaluators know they are AI-generated

Interestingly, this finding is especially noteworthy when compared to Magni et al. [2024]'s *Humans as Creativity Gatekeepers: Are We Biased Against AI Creativity?*. Magni et al. [2024]'s findings show that people often rated AI-outputs as less creative, when told that they are AI-generated. However, in our second study, Study 2, the participants who evaluated the AI-assisted and not AI-assisted outputs were not informed of the AI condition at all. While the self-ratings for creativity shown in previous sections demonstrate lower ratings for the AI-assisted content. This could be an indicator of producer identity bias [Magni et al., 2024], where people rate AI-generated or AI-assisted results as less creative, when they are aware of the contents' origin.

AI-assisted designs ranked higher for clarity, innovation, and practicality, emphasizing a balance between originality and usability

In the survey, participants were also asked to rank the products from best to worst, without being informed at all about the AI and No AI conditions. The participants reported ranking the products (the Smart Plant Pot and the Glow-in-the-Dark Umbrella) based on creativity, practicality, and innovation. The highest-ranked umbrella design was AI-assisted, and the second highest was created without AI. Participants justified their rankings based on simplicity, personalization, and eco-friendliness. Regarding the smart plant pot, the highest two ranked designs were AI-assisted, and the participants reported ranking based on novelty, unique features (i.e. solar power, connectivity via app), and memorable features. Designs that were perceived as confusing or having less features ranked lower. To conclude, AI-assisted creative outputs were generally preferred overall, which suggests that AI has the ability to enhance clarity, innovative nature, and overall appeal in creative scenarios. However, participant feedback also shows that it is important to achieve a balance between originality and usability and ensure ideas are clear and interesting.

Self-ratings favored practicality and structure, while external ratings emphasized originality and unpredictability

In Section 5.2, we also compared the self-ratings with the

external-ratings for the creative outputs in each condition. Interestingly, the pattern of ratings is similar across both AI and No AI conditions. Participants generally rated their own ideas higher in practicality, organization, definition and appeal compared to external evaluations. However, external ratings were higher for originality and unpredictability, both in the AI and No AI conditions. This highlights how creators may focus more on the practical and structured qualities of their work, while external observers tend to focus on original and unpredictable aspects.

6.2.3 Task Efficiency and Word Count

Transcripts allowed calculation of w.p.m.

During the analysis of the videos from Study 1, it was possible to find the word count for each group in both conditions using the transcripts we had generated. This allowed us to calculate the task efficiency in each condition, as well as calculate the words per minute in each condition.

First tasks took longer due to learning effects, aligning with the Instance Theory of Automaticity

Our first calculation shows that in all cases, the first condition took longer than the second condition, regardless of the condition itself. This can be explained by the Instance Theory of Automaticity, according to Logan [1992]. The theory suggests that reaction times decrease as individuals accumulate traces of memory upon repeating a task, which leads to faster performance. This explains why groups across the entire study were slower during their first task regardless of the condition, since they were at the initial learning phase [Logan, 1992].

Adjusted for learning effects, most groups took longer in the AI condition

Therefore, to analyze the time differences between conditions, we calculated the mean difference between conditions across all groups and added it to the second task's time. This way we could assess if the difference between conditions is significant. Our calculations resulted in a mean difference of 7 minutes, which was added to the second task time in each group. This resulted in four groups having a **higher** task-completion time during the **AI condition**, and the two remaining groups having nearly the same time for both conditions.

The findings suggest that the AI condition was generally slower. This is likely due to the additional time that multiple groups spent generating images for the drawing task, which often took multiple minutes. This could possibly explain the overall increase in task completion time in the AI condition.

AI condition was slower, likely due to time spent generating images

In addition to measuring the time needed for each task, we measured the number of words spoken in each condition, and consequently the words per minute (w.p.m) in each group for both conditions. For context, the average active conversation in English has 210 words per minute [Tauroza and Allison, 1990]. According to Pimsleur et al. [1977], a speech rate below 130 w.p.m is categorized as *slow*.

Speech rates in each condition were analyzed

The measured words-per-minute were higher in the No AI condition (M = 100, SD = 13) than in the AI condition (M = 90, SD = 13). On average, ten more words-per-minute were spoken in the No AI condition, indicating that participants engaged in more verbal communication when there was no available AI. This could be explained by the increased brainstorming, discussion, and collaborative ideation that was observed in the qualitative discussion. Moreover, according to the categorization of Pimsleur et al. [1977], the speech rate during both conditions is considered slow. This could be due to the nature of the study. The participants took the time to discuss their options, look at a screen and read information and also pause to think about their decisions, which led to longer pauses and smaller speech rates.

No AI condition had a higher speech rate (w.p.m.

6.3 Limitations

In Study 1, we had groups of three work together on creative tasks in marketing, once with AI assistance, and once without. After reporting the results and discussion, it is also important to note the limitations of Study 1. Among the most important limitations is that the qualitative analysis of Study 1 was performed by a single coder, meaning there is reduced inter-coder reliability. This is one of the main reasons that Reflexive Thematic Analysis was implemented, rather than Thematic Analysis. In addition, the

Study 1 limitations include single-coder qualitative analysis and a small sample

small sample size of 18 participants during Study 1, among whom many participants have a technical or scientific background, could potentially limit the generalizability of the results.

Study 2 limitations include potential bias toward aesthetics and residual order effects in evaluations

In Study 2, we conducted a survey for external participants to rate and rank the creative results from Study 1, a possible limitation could be the bias towards aesthetics over substance regarding the creative results. In addition, although the the order of the creative outputs in Study 2 was randomized, it does not entirely eliminate the possibility of an order effect. Participants may still have been influenced by creative results they had seen before, leading to potential biases in their evaluations, therefore affecting ranking and rating outcomes.

Chapter 7

Summary and Future Work

In this chapter, we conclude the thesis by summarizing the process and outcomes of the thesis and its studies, and afterwards suggest future research directions based on our findings.

7.1 Summary and Contributions

For this thesis, we aimed to explore how *generative AI tools impact the creative process and outcomes in co-located group processes*, in which we initially conducted a lab study with 18 participants split into six groups, followed by a second survey study. During **Study 1**, the groups of three worked together on a marketing task for a given product, consisting of four sub-tasks: creating a product name, slogan, selling points and drawing the product with its logo. The participants were exposed to two conditions, the AI condition and the no AI condition. The tasks were identical in both conditions, with the only exception being the product that is marketed. Using Reflexive Thematic Analysis [Braun and Clarke, 2019], we evaluated the difference between the group behaviors during these two conditions. In addition, we evaluated the results from the questionnaires that were

Thesis explores how generative AI impacts creative processes in co-located groups through two studies

given to the participants of the user study. Afterwards, in **Study 2**, we allowed external participants to review and evaluate the creative outputs of Study 1. These two studies resulted in interesting findings that addressed our research questions.

Our findings revealed that generative AI tools had a notable impact on group creativity, collaboration dynamics, and perceived creative outcomes. In the first study, we observed that groups using AI interacted differently compared to those in the No AI condition. AI-assisted groups often engaged in quicker idea generation by leveraging AI-generated ideas for inspiration. However, they also displayed a tendency to rely on AI's outputs rather than engaging in deeper discussions to refine ideas collaboratively. In contrast, when the groups worked without AI, they participated in more extended discussions and negotiations, leading to a more organic creative process. The questionnaire analysis for the first study provided further insights into participants' perceptions of their creativity and group collaboration across conditions. While some participants found AI tools helpful in reducing cognitive effort and sparking new ideas, others felt that AI reduced their creative autonomy by influencing their ability to make decisions. Additionally, self-perceived creativity ratings showed that participants tended to rate their work as more original in the no-AI condition. In **Study 2**, external evaluators assessed the creative outputs from both conditions, allowing us to examine how AI-assisted creativity is perceived from an outside perspective. Interestingly, the external creativity ratings were generally higher for the AI-assisted creative outputs, despite the external raters having not been informed about two conditions from Study 1 (AI and No AI). Another interesting finding is that the pattern of ratings was the same across both conditions, with participants generally rating their own ideas higher in practicality, organization, definition and appeal, while external evaluators gave higher ratings for originality and unpredictability. This pattern was identical in both AI and No AI conditions.

AI influenced creativity accelerating idea generation but reduced deep discussion. Self-ratings of creativity favored No AI, while external ratings favored AI outputs

7.2 Future Work

In summary, our research provides valuable insights into how AI tools influence group creativity. Although AI can serve as a useful support tool in creative collaboration, its adoption in creative group settings should be carefully considered by making it enhance group creativity and collaboration, rather than overpower it. Our findings contribute to the research gap on group collaboration and AI, particularly in co-located group settings, which is relevant in both educational and professional settings with the rapidly increasing adoption of AI.

Findings address the research gap on AI in co-located group settings

For future studies, we propose conducting **longitudinal studies** to explore how group creativity evolves when exposed to AI over time. While our study involved single sessions with AI, exposure over several months or more could produce new patterns of collaboration, dependency, or skill growth. This could lead to a deeper understanding on whether prolonged AI usage helps or diminishes group creativity. Furthermore, tracking changes in behavior across several creative sessions may also reveal if participants grow more skilled and productive at using AI or if they develop over-reliance on AI.

Future studies could explore long-term AI exposure

In addition, future studies can explore **adaptive AI systems** that respond dynamically to changing participant behavior and requirements. For example, an adaptive AI would begin with minimal input and gradually add guidance depending on the participant's needs. These systems could be designed to enhance human creativity while preventing over-reliance on AI. On a broader scale, future work could explore how AI influences **different creative and cultural settings**. While our study was focused on marketing tasks, AI generative models are applied increasingly in the field of visual arts, music composition, and industrial design. Therefore, conducting studies on how AI is used across a number of creative fields could provide a larger understanding of its impact.

Future studies could investigate adaptive AI systems

By building upon our findings and expanding research into broader areas, future research can contribute to building

AI tools that enhance group creativity, while also addressing social, ethical, and cognitive challenges that accompany human-AI interaction.

Appendix A

Questionnaires and Documents

Consent Form for Participation in a Research Study

Principal Investigator: Raghad Zaghal
Institution: Media Computing Group (Chair i10) at RWTH Aachen

Purpose of the Study

The purpose of this study is to explore how generative AI tools impact the creative process in groups.

Procedures

If you agree to participate:

1. You will take part in an approx. 60-minute session involving creative tasks in a group of three participants.
2. Your results and interactions will be video recorded for analysis. The video recordings will not be published.

Confidentiality

- All data collected during this study will remain confidential.
- Your responses will be anonymized and stored securely, all personal data and recordings will be deleted after completing the research project.
- Any published findings will not include information that can identify you or your group.

Voluntary Participation

Your participation is voluntary. You may decline to answer any question or withdraw from the study at any time without explanation or penalty.

Contact Information

If you have questions about the study, you may contact Raghad Zaghal at raghad.zaghal@rwth-aachen.de.

Participant Consent

Please read and sign below to indicate your agreement to participate.

- I have read and understood the information provided in this consent form.
- I understand that my participation is voluntary and that I may withdraw at any time without penalty.
- I agree to allow my data (task outputs, responses, and interactions) to be used for research purposes under the conditions described.
- I understand that my identity will remain confidential.

Participant Name: _____

Participant Signature: _____ **Date:**_____

Study ID
Participant

Background Questionnaire

Demographic

1. **Age**:

2. **Gender**:

   ☐ Male ☐ Female

   ☐ Non-binary/Other ☐ Prefer not to say

3. **Highest Level of Education or Current Program**

   ☐ High school diploma ☐ Bachelor's degree

   ☐ Master's degree ☐ Doctoral degree

   ☐ Other (please specify):

4. **Field of Study/Profession**:

Creativity Background

5. **How often do you engage in creative activities?**

   Creative activities include tasks such as drawing, painting, writing, designing, crafting, or brainstorming new ideas.

   ☐ Never ☐ Once a month ☐ Once a week

   ☐ 2-3 Times/ week ☐ Daily

6. **Do you have experience working on creative tasks in a group setting? If yes, what?**

   ☐ Yes, ☐ No

7. **How would you evaluate your creativity in problem-solving and idea generation? (Mark only one option)**

Rarely think of creative solutions	Occasionally come up with creative ideas	Often think creatively in familiar situations	Frequently generate innovative ideas in diverse contexts	Consistently develop unique and impactful ideas
○	○	○	○	○

Familiarity with AI

8. **Have you used generative AI tools before (e.g., ChatGPT, MidJourney, DALL·E)? If yes, what?**

☐ Yes, _____ ☐ No

9. **If yes, how often do you use these tools?**

☐ Daily ☐ Weekly ☐ Monthly ☐ Rarely

10. **For what purposes have you used generative AI tools?** (Select all that apply)

☐ Work-related tasks ☐ Academic purposes

☐ Personal projects ☐ Entertainment

☐ Other (please specify): _____

11. **How would you rate your familiarity with AI tools? (Mark only one option)**

No experience with AI	Minimal experience with AI	Some experience with AI	Considerable experience with AI	Extensive experience with AI
○	○	○	○	○

Group Work

12. **How well do you know your group members?**

Not at all	Slightly	Moderately	Quite a lot	Very much
○	○	○	○	○

13. **How much do you enjoy working in a group?**

Not at all	Slightly	Moderately	Quite a lot	Very much
○	○	○	○	○

14. **Do you prefer leading group discussions or contributing equally with others?**

- ☐ Prefer leading roles
- ☐ Prefer equal contribution
- ☐ Prefer secondary roles
- ☐ No preference

Marketing Task Instructions

Product: Glow-in-the-Dark Umbrella

Your group is tasked with creating a **marketing** strategy for the assigned product. Collaborate to come up with creative and impactful ideas to showcase the product effectively. The details are up to you!

The product being marketed is a: Glow-in-the-Dark Umbrella

Subtasks:

For each subtask, brainstorm as a group and record your results in the worksheet provided.

1. **Come up with a name for the product**
2. **Create a catchy slogan**
3. **List 2-3 key selling points**
4. **Draw the product <u>including</u> a logo**

Notes:

- **Materials Provided:** Colored pens, blank paper, handouts, and a worksheet.
- **Worksheet:** Please only use the group worksheet for writing your answers.

Study ID

Condition

Product

Marketing Worksheet

1. Come up with a name for the product

2. Create a catchy slogan

3. List 2-3 key selling points

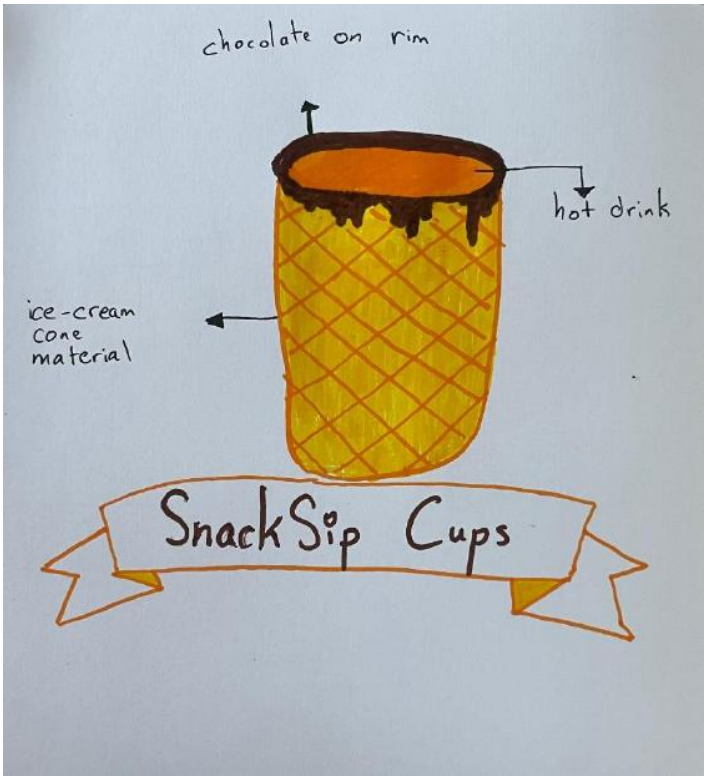
4. Draw the product including a logo

Marketing Task Example

Product: Edible Cup

Subtasks with examples for guidance.

1. **Come up with a name for the product**
    - *Example:* SnackSip Cups
2. **Create a catchy slogan**
    - *Example:* "Sip, Snack, and Save the Planet"
3. **List 2-3 key selling points**
    - *Example:* 100% biodegradable, comes in different flavors, durable for hot and cold beverages
4. **Draw the product including a logo!**
    - *Example:*



Study ID ☐ Condition ☐
Participant ☐ Product ☐

Questionnaire After Each Task

1. Collaboration Dynamics

Instructions: Please evaluate the collaboration dynamics within your group after completing the task. For each statement, write the number that best represents your level of agreement, where:

Scale: Strongly Disagree - Strongly Agree

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1.1. Our group has open communication.	O	O	O	O	O
1.2. Our team members understand what our task is.	O	O	O	O	O
1.3. Our team members help one another deal with problems or resolve issues.	O	O	O	O	O
1.4. We are able to work through differences of opinion without causing conflict.	O	O	O	O	O

2. Task-Specific Creativity Perception

Instructions: Please rate your group's **marketing task result** by marking **only one** number on each row, where -2 indicates the left-hand adjective and 2 indicates the right-hand adjective.

Scale: -2 to 2.

	-2	-1	0	1	2	
2.1. Conventional	O	O	O	O	O	Original
2.2. Predictable	O	O	O	O	O	Unexpected
2.3. Unfocused	O	O	O	O	O	Well-defined
2.4. Impractical	O	O	O	O	O	Practical
2.5. Unappealing	O	O	O	O	O	Appealing
2.6. Disorganized	O	O	O	O	O	Organized

3. Emotional and Cognitive Load

Instructions: Please rate each question by writing a number from **Very Low** to **Very High** in the box provided.

Scale: Very Low – Very High

Statement	Very Low	Low	Neutral	High	Very High
3.1. How much mental and cognitive effort was required to complete this task?	○	○	○	○	○
3.2. How much time pressure did you feel to complete the task?	○	○	○	○	○
3.3. How satisfied are you with your performance on this task?	○	○	○	○	○
3.4. How much effort did you need to put into completing the task?	○	○	○	○	○

4. Digital Tool's Role and Perceived Usefulness

Instructions: In this condition, your group **used** the **digital tool (ChatGPT)** to complete the task. Please rate your experience by choosing **one** answer for each question provided.

Scale: Strongly Disagree - Strongly Agree

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
4.1. I felt inspired.	○	○	○	○	○
4.2. This task was enjoyable.	○	○	○	○	○
4.3. Our team collaborated well.	○	○	○	○	○
4.4. I appreciated the creative input of my team.	○	○	○	○	○
4.5. I needed more help.	○	○	○	○	○
4.6. Our team worked too slowly on our task.	○	○	○	○	○

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
4.7. I could think independently.	○	○	○	○	○
4.8. I felt distracted by the digital tool.	○	○	○	○	○
4.9. I was confident about when to use the digital tool.	○	○	○	○	○
4.10. The digital tool limited my creativity.	○	○	○	○	○

5. Open-Ended Feedback

This Condition = Using the Digital Tool

5.1. What were your overall thoughts or impressions about working in **this condition**?

5.2. What aspects of **this condition** did you find most helpful, if any?

5.3. What challenges did you face while working under **this condition**, if any?

5.4. In what ways did you use the **digital tool**, and for what purpose? (e.g., generating ideas, refining ideas, etc.)

Study ID
Participant

Final Questionnaire

1. Which product did you overall prefer marketing?

☐ Glow-in-the-Dark Umbrella ☐ Smart Plant Pot ☐ No preference

Please justify your answer:

2. Which condition did you overall prefer for completing the tasks?

☐ Not using AI ☐ Using AI ☐ No preference

Please justify your answer:

3. During which condition do you feel your group produce the most creative results?

☐ 1st condition ☐ 2nd condition ☐ Both equally

4. During which condition do you feel your group worked best together?

☐ 1st condition ☐ 2nd condition ☐ Both equally

5. Did the AI feel more like a teammate, a guide, or a leader during the tasks? Why?

☐ Teammate ☐ Guide ☐ Leader

☐ Other. Please explain:

6. (Optional) How would you briefly describe your overall experience working on the tasks?

User Study Numbering System

Study ID

01 – 02 – 03 – 04 – 05 – 06

Participant

X (left seat) – **Y** (middle seat) – **Z** (right seat).

Condition

A (AI) – **N** (No AI)

Product

U (Umbrella) – **P** (Plant Pot)

Document Abbreviations

C	Consent Form
B	Background Questionnaire
H	Handout
W	Worksheet
S	Scribble Sheet
Qa	After Task Questionnaire (AI Condition)
Qn	After Task Questionnaire (No AI Condition)
F	Final Questionnaire

Examples

02_Y_ Qn_N_P = Group 02, User Y (Middle Seat), After Task Questionnaire (No AI condition), No AI, Plant Pot as a product.

04_Z_ Qa_A_U_2.3 = Group 04, User Z (Right Seat), After Task Questionnaire (AI condition), AI, Umbrella as a product, Question 2.3.

05_X_B_4 = Group 05, User X (Left Seat), Background Questionnaire, Question 4.

Order

Group	User (if applicable)	Document	Condition (if applicable)	Product (if applicable)	Question (if applicable)

Study 2

Informed Consent Form

Thank you for participating in this study. Before we begin, please read the following information carefully. Your participation is voluntary, and you may withdraw from the study at any time without providing a reason.

Data Protection and Confidentiality

All data collected will be anonymized and analyzed in aggregate form. No personally identifiable information will be recorded. The data will be used solely for academic research purposes.

Voluntary Participation

Your participation is entirely voluntary. You may withdraw at any time without any consequences. There are no known risks associated with participating in this study.

Contact Information

If you have any questions about this study, please feel free to contact the researchers:

Researcher: Raghad Zaghal
Institution: RWTH Aachen University
Email: raghad.zaghal@rwth-aachen.de

Consent Statement

By clicking "Next", you confirm that:

- You have read and understood the information provided.
- You voluntarily agree to participate in this study.
- You are at least 18 years old.

If you do not wish to participate, you may close this window.

Thank you for participating in this survey. In a previous study, groups of three collaborated on creative tasks. This survey aims to gather your insights on the results of that study.

You will be shown different creative results from the previous study and asked to evaluate them based on specific criteria. There are no right or wrong answers—please provide your honest opinions. The survey should take approximately 15 minutes to complete.

Let's get started!

1. Task Explanation

In the previous study, participants worked in groups of three to complete a marketing task. Each group was assigned the same product and was responsible for developing a creative marketing concept for it. To structure their approach, participants were given four specific subtasks:

1. Define the product **name**.
2. Develop a catchy **slogan**.
3. **Sketch the** product and its logo.
4. Highlight 2-3 key **selling points**.

Each group completed this task for two products.
In this survey, you will **evaluate the creative results** generated during these sessions.

Note: Please rate based on idea quality, not execution or drawing skill.

◯ I have read and understood my task.

Next

Given Product Type: Smart Plant Pot

Product Name:	Plant Pal
Product Slogan:	"Grow smarter, not harder"
Selling Points:	1. Ideal for budget conscious and beginner plant parents. 2. App notifications – never forget watering your plants again. 3. Comes with sensors for all kinds of plants.
Product and Logo Drawing	
	

Conventional	○	○	○	○	○	Original
Predictable	○	○	○	○	○	Unexpected
Unfocused	○	○	○	○	○	Well-defined
Impractical	○	○	○	○	○	Practical
Unappealing	○	○	○	○	○	Appealing
Disorganized	○	○	○	○	○	Organized

Given Product Type: Smart Plant Pot

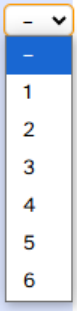
Product Name:	BotaniQ
Product Slogan:	"BotaniQ: The Smart Way to a Greener Thumb!"
Selling Points:	1. Controllable with an App. 2. Registers overwatering/dryness/sun exposure (with a self-watering setting). 3. Solar-powered.
Product and Logo Drawing	
	

. . . The **rating** continues for all 6 Smart Plant Pot creative results.

2. Ranking the Plant Pots

Rank the Smart Plant Pot marketing results from best (1) to worst (6).
Please rank based on **idea quality**, not execution or drawing skill.
Please rank each set of answers **(text and picture)** together.

Product Name:	Smartpot
Product Slogan:	"Be sprouty"
Selling Points:	1. Powered by solar power. 2. Manageable by app. 3. Individual/customized plan for plant.
Product and Logo Drawing	



. . . The **ranking** continues for all 6 Smart Plant Pot creative results.

3. Justify Ranking (Smart Plant Pot)

Please briefly justify your ranking.

Next

After ranking the results, a brief textual **justification** is required.

Given Product Type: Glow-in-the-Dark Umbrella

Product Name:	Lumibrella
Product Slogan:	"Rain won't dull you when you shine"
Selling Points:	1. Safer night time walk. 2. It lights you up in bad weather. 3. Eco-friendly glow technology.
Product and Logo Drawing	

Conventional	○	○	○	○	○	Original
Predictable	○	○	○	○	○	Unexpected
Unfocused	○	○	○	○	○	Well-defined

. . . The **rating** continues for all 6 Glow-in-the-Dark Umbrella creative results.

4. Ranking the Umbrellas

Rank the Glow-in-the-Dark Umbrella marketing results from best (1) to worst (6).
Please rank based on idea quality, not execution or drawing skill.

Product Name:	Luminella
Product Slogan:	"See and be seen!"
Selling Points:	1. Fashionable accessory (fun color options). 2. Increased traffic safety for pedestrians.
Product and Logo Drawing	

The **ranking** continues for all 6 Glow-in-the-Dark Umbrella creative results.

After ranking the results, a brief textual **justification** is required.

END OF SURVEY

Appendix B

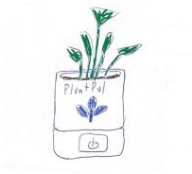
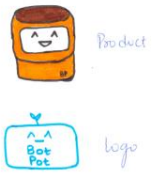
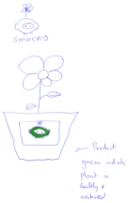
First User Study: Creative Outputs

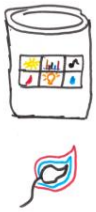
Glow-in-the-Dark Umbrella (Creative Output)

Group 01	Group 02	Group 03
Funbrella	Lumibrella	Starella
Hey little fella, grab a Funbrella!	Rain won't dull you when you shine	When Rain Meets Radiance
1. Safety for kids on the way to school and back. 2. Fun colors and designs; your kid will remember to take it! 3. Eco-friendly!	1. Safer night time walk. 2. It lights you up in bad weather. 3. Eco-friendly glow technology.	1. Safety. 2. Personlization. 3.Solar - energy use => Eco-friendly product
		

Group 04	Group 05	Group 06
Luminella	Starella	Lumbrella
See and be seen!	Let the night shine on you!	Just Glow
1. Fashionable accessory (fun color options). 2. Increased traffic safety for pedestrians.	1. Custom glow designs: Choose from multiple personalized glowing patterns. 2. Solar powered glow: Self-sustained eco-friendly power charging. 3. Rain activated sensors: Intensify the glow during rain, creating dazzling, practical light displays.	1. Lightweight. 2. Stormproof. 3. Luminescent.
		

Smart Plant Pot (Creative Output)

Group 01	Group 02	Group 03
Plant Pal	BotPot	Smacky
Grow smarter, not harder	Your green-thumbed pot	Smacky - complement your green friends!
1. Ideal for budget conscious and beginner plant parents. 2. App notifications - never forget watering your plants again. 3. Comes with sensors for all kinds of plants.	1. Cute digital plant pot. 2. No expertise needed. 3. Track your plant's health.	1. Smart - takes care if you forget. 2. Suitable for every plant. 3. Saves water and energy.
		

Group 04	Group 05	Group 06
BotaniQ	Piezo Plant Pot	Smartpot
BotaniQ: The Smart Way to a Greener Thumb!	The sound of nature - in your home.	Be sprouty
1. Controllable with an App. 2. Registers overwatering/dryness/sun exposure (with a self-watering setting). 3. Solar-powered.	1. Ai drived irrigation system. 2. Real-time ambient sound generation using piezo sensor technology. 3. Customizable sound experience.	1. Powered by solar power. 2. Managable by app. 3. Individual/customized plan for plant.
		

Bibliography

[1] Nagadivya Balasubramaniam, Marjo Kauppinen, Kari Hiekkanen, and Sari Kujala. Transparency and explainability of AI systems: ethical guidelines in practice. In *International working conference on requirements engineering: foundation for software quality*, pages 3–18. Springer, 2022.

[2] Ajay Bandi, Pydi Venkata Satya Ramesh Adapa, and Yudu Eswar Vinay Pratap Kumar Kuchi. The power of generative ai: A review of requirements, models, input–output formats, evaluation metrics, and challenges. *Future Internet*, 15(8):260, 2023.

[3] Susan P Besemer and Karen O'Quin. Analyzing creative products: Refinement and test of a judging instrument. *The Journal of Creative Behavior*, 1986.

[4] Katerina Bezrukova, Terri L Griffith, Chester Spell, Vincent Rice Jr, and Huiru Evangeline Yang. Artificial intelligence and groups: Effects of attitudes and discretion on collaboration. *Group & Organization Management*, 48(2):629–670, 2023.

[5] Marvin Braun, Maike Greve, and Ulrich Gnewuch. The New Dream Team? A Review of Human-AI Collaboration Research From a Human Teamwork Perspective. In R. De', editor, *Rising like a Phoenix: Emerging from the Pandemic and Reshaping Human Endeavors with Digital Technologies ICIS 2023*, ICIS 2023 Proceedings. 44th International Conference on Information Systems (ICIS 2023), AIS eLibrary (AISeL), 2023.

[6] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[7] Virginia Braun and Victoria Clarke. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health*, 11(4):589–597, 2019.

[8] Vincent Brown, Michael Tumeo, Timothy S Larey, and Paul B Paulus. Modeling cognitive interactions during group brainstorming. *Small group research*, 29(4):495–526, 1998.

[9] Gary Charness, Uri Gneezy, and Michael A. Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1):1–8, 2012. doi.org/10.1016/j.jebo.2011.08.009.

[10] Venita DePuy and Vance W Berger. Counterbalancing. *Wiley StatsRef: Statistics Reference Online*, 2014.

[11] Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch, and Patrick Zschech. Generative ai. *Business & Information Systems Engineering*, 66(1):111–126, 2024.

[12] Jennifer M George. Extrinsic and intrinsic origins of perceived social loafing in organizations. *Academy of Management Journal*, 35(1):191–202, 1992.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[14] Jean-Christophe Goulet-Pelletier and Denis Cousineau. A review of effect sizes and their confidence intervals, Part I: The Cohen'sd family. *The Quantitative Methods for Psychology*, 14(4):242–265, 2018.

[15] Sandra G. Hart. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 50, pages 904–908. Sage Publications, 2006.

[16] Angel Hsing-Chi Hwang. Too late to be creative? AI-empowered tools in creative processes. In *CHI conference on human factors in computing systems extended abstracts*, pages 1–9, 2022.

[17] Chipp Jansen and Elizabeth Sklar. Exploring co-creative drawing workflows. *Frontiers in Robotics and AI*, 8:577770, 2021.

[18] Han Kyul Kim, Aleyeh Roknaldin, Shriniwas Prakash Nayak, Xiaoci Zhang, Muyao Yang, Marlon Twyman, Angel Hsing-Chi Hwang, and Stephen Lu. ChatGPT and Me: Collaborative Creativity in a Group Brainstorming with Generative AI. In *2024 ASEE Annual Conference & Exposition*, 2024.

[19] Chunyan Liu, Zhe Ren, and Sen Liu. Using design and graphic design with color research in AI visual media to convey. *Journal of Sensors*, 2021(1):8153783, 2021.

[20] Gordon D Logan. Shapes of reaction-time distributions and shapes of learning curves: a test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5):883, 1992.

[21] Yingyue Luna Luan. *The New Creative Alliance: Investigating the Dynamics of Human-AI Collaboration in Creative Endeavours*. PhD thesis, University of Cambridge, 2024.

[22] Federico Magni, Jiyoung Park, and Melody Manchi Chao. Humans as creativity gatekeepers: Are we biased against AI creativity? *Journal of Business and Psychology*, 39(3):643–656, 2024.

[23] Birgit Muskat, Amitabh Anand, Christine Contessotto, Adrian Heng Tsai Tan, and Guihyun Park. Team familiarity—Boon for routines, bane for innovation? A review and future research agenda. *Human Resource Management Review*, 32 (4):100892, 2022.

[24] Bernard A. Nijstad. Creativity in groups. In Mario Mikulincer, Phillip R. Shaver, John F. Dovidio, and Jeffry A. Simpson, editors, *APA Handbook of Personality and Social Psychology, Vol. 2: Group Processes*, pages 35–65. American Psychological Association, 2015. doi.org/10.1037/14342-002.

[25] Bernard A Nijstad and Wolfgang Stroebe. How the group affects the mind: A cognitive model of idea generation in groups. *Personality and social psychology review*, 10(3):186–213, 2006.

[26] Shakked Noy and Whitney Zhang. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192, 2023.

[27] Norvig Peter and Russell Stuart Artificial Intelligence. *A Modern Approach*. Pearson Education, USA, 2021.

[28] Paul Pimsleur, Charles Hancock, and Patricia Furey. Speech rate and listening comprehension. *Viewpoints on English as a second language*, 27, 1977.

[29] Stanley Presser, Mick P Couper, Judith T Lessler, Elizabeth Martin, Jean Martin, Jennifer M Rothgeb, and Eleanor Singer. Methods for testing and evaluating survey questions. *Methods for testing and evaluating survey questionnaires*, pages 1–22, 2004.

[30] Priscilla M Pyett. Validation of qualitative research in the "real world". *Qualitative health research*, 13(8):1170–1179, 2003.

[31] Raimund Schatz, Sebastian Egger, and Kathrin Masuch. The impact of test duration on user fatigue and reliability of subjective quality ratings. *Journal of the Audio Engineering Society*, 60(1/2):63–73, 2012.

[32] Christian TK-H Stadtländer. Qualitative, quantitative, and mixed-methods research. *Microbe Magazine (3rd ed., Vol. 4, Issue 11). Sage Publications. https://doi.org/10.1128/microbe*, 4(1), 2009.

[33] Steve Tauroza and Desmond Allison. Speech rates in british english. *Applied linguistics*, 11(1):90–105, 1990.

[34] Ivana Tomić, I Juric, Sandra Dedijer, and Savka Adamović. Artificial Intelligence in Graphic Design. In *Proceedings of the 54th Annual Scientific Conference of the International Circle of Educational Institutes of Graphic-Media Technology and Management, The Hellenic Union of Graphic Arts and Media Technology Engineers, Greece*, pages 85–93, 2023.

[35] Kunfeng Wang, Chao Gou, Yanjie Duan, Yilun Lin, Xinhu Zheng, and Fei-Yue Wang. Generative adversarial networks: introduction and outlook. *IEEE/CAA Journal of Automatica Sinica*, 4(4):588–598, 2017. doi.org/10.1109/JAS.2017.7510583.

[36] Liang Yu. Ethical considerations in case studies. In *2020 1st Workshop on Ethics in Requirements Engineering Research and Practice (REthics)*, pages 15–21. IEEE, 2020.

[37] I Made Marthana Yusa, Yu Yu, and Tetiana Sovhyra. Reflections on the use of artificial intelligence in works of art. *Journal of Aesthetics, Design, and Art Management*, 2(2):152–167, 2022.

[38] Zheng Zhang, Weirui Peng, Xinyue Chen, Luke Cao, and Toby Jia-Jun Li. LADICA: A Large Shared Display Interface for Generative AI Cognitive Assistance in Co-Located Team Collaboration. *arXiv preprint arXiv:2409.1 3968*, 2024.

Index

Typeset March 24, 2025