# IIA4XAI: Interactive Installation Art to Increase Non-Expert User Engagement with AI Systems

**Graduation Thesis, June 2023**

Marise van Noordenne[1]
Advisors: Dr. Rob Saunders[1] and Dr. Peter van der Putten[1]

MSc Media Technology
[1] Leiden Institute of Advanced Computer Science, Leiden University
Niels Bohrweg 1, 2333 CA Leiden
marisevannoordenne@gmail.com

**Abstract.** The rise of consumer-focused Artificial Intelligence (AI) applications like ChatGPT, DALL-E 2, Stable Diffusion, and Jasper AI has made these technologies more accessible to the public. Yet, despite using these applications regularly, most people still find the underlying technology complex and mysterious and lack a clear understanding of how these systems work. Explainable Artificial Intelligence (XAI) aims to address this issue by developing explainable models. However, current XAI systems primarily cater to technical users with prior knowledge of the field, leaving non-expert users empty-handed. The current study investigates the introduction of Interactive Installation Art (IIA) into the design of XAI systems as a prospective solution to enhance initial user engagement and increase accessibility for non-expert users, by allowing them to interact with the properties of the system in a playful and informal manner. The study presents the IIA for XAI (IIA4XAI) framework that serves as a foundation for utilizing IIA as a design technique in the development of XAI systems. It offers valuable guidance to designers and engineers who aim to involve non-expert users as part of their target audience. The framework is based on an exploration of XAI and IIA theories and a preliminary study conducted through a survey. To evaluate the framework, a case study was performed, developing an interactive art installation built around AI image generation technology. Qualitative user tests were conducted on the installation prototype, demonstrating the effectiveness of the IIA4XAI framework in engaging non-expert users with AI systems. Overall, this study highlights the potential of the IIA4XAI framework as a design technique to make AI systems more understandable and accessible to a broader audience of non-expert users.

**Keywords:** Explainable Artificial Intelligence · Explainable Computational Creativity · Human Computer Interaction · Interactive Installation Art

## 1 Introduction

The recent surge in consumer-focused Artificial Intelligence (AI) applications like ChatGPT, DALL-E 2, Stable Diffusion, and Jasper AI, sparked attention from a broad audience towards Generative AI (GenAI). Where previously, GenAI and Machine Learning (ML) applications were limited to academic research and big tech companies, they have now made their way into the daily lives of individuals and businesses. AI technologies have already been common practice in consumer applications, for example, in recommendation algorithms used by Netflix and Spotify, Facebook's friend suggestions, or AI-powered virtual assistants like Siri, Alexa, and Google Assistant. But applications such as ChatGPT and DALL-E 2 have shifted the role of AI from being in the background to enabling users to actively interact with the system for targeted tasks.

Despite the popularity of GenAI, most people still find the technology behind these applications complex and incomprehensible and have little to no understanding of the actual inner workings of the system. This lack of knowledge could lead to potential dangers such as the spread of false beliefs, adaptation of biases, or over-reliance on the system. The field of Explainable Artificial Intelligence (XAI) [1] seeks to address this issue by developing systems that provide users with insights into the model's values and decision-making processes. Current XAI systems, however, are primarily targeted at technical users with pre-existing knowledge of the field, which means that non-expert users, who could benefit the most from these systems, are left with no tools to help them understand.

The study presented here investigates the use of aesthetics, in the form of Interactive Installation Art (IIA), in the design of XAI systems as a prospective solution to enhance user engagement and increase accessibility for non-expert users. The introduction of an interactive art installation allows users to initially come into contact with the workings and properties of the AI system in a playful and informal manner. This allows users to observe and experiment with the technology, fostering an interest in learning about and understanding its inner workings, and creating engagement. As a result, the aim of this study is to develop a framework for the utilization of interactive installation art as a technique for creating user engagement with the technology as a first step in implementing XAI. Specifically, this study will focus on AI image generation technologies as a case study to evaluate and examine the framework.

This thesis starts with an exploration of related work in the fields of XAI, XCC, and IIA in Section 2. A quantitative preliminary study was conducted to gain insights into the current attitudes and understanding of AI systems and the potential of the introduction of IIA for increasing engagement with non-expert users. The details and results of this preliminary study are presented in Section 3. The outcomes of the exploration of related work and the preliminary study were used for the development of the IIA4XAI framework, which is introduced and elaborated upon in Section 4. Finally, Section 5 describes the execution and results of the case study that involved the development and qualitative user testing of an interactive art installation based on AI image generation, as a method to evaluate the IIA4XAI framework. In Figure 1 a schematic overview of the current study is presented.

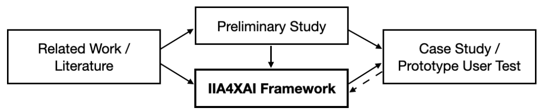


**Fig. 1.** Schematic overview of the study design

## 2 Related Work

The following section explores the intersection of XAI and IIA. XAI addresses the limitations of inherently uninterpretable AI models by developing more transparent models with the goal of increasing trust and understanding in these systems. Additionally, the field of Explainable Computational Creativity (XCC) is explored. XCC focuses on enabling collaboration and communication between humans and computational creative systems, with a focus on co-creativity. The principles of XCC differ from those of XAI by putting more emphasis on intuitive communication and accessibility for human users. Finally, the domain of IIA is introduced. This artistic domain is known for its engagement with recipients and integration of new technologies. By removing barriers and creating a non-intimidating, playful environment, interactive installations facilitate hands-on exploration of technology and foster curiosity.

### 2.1 Explainable Artificial Intelligence (XAI)

In the field of AI, there are generally two different types of models: inherently interpretable and inherently uninterpretable AI models [2]. Inherently interpretable AI models are Machine Learning (ML) algorithms like decision trees or Bayesian classifiers that consist of model components, like a path in a decision tree, that can be directly inspected and used to gain an understanding of the model's predictions. This makes the model inherently interpretable and provides traceability and transparency in the decision-making process. This type of model is also called a "white box model" [3]. On the other hand, there are Deep Learning (DL) algorithms that generate more complex models and operate in unintuitive dimensions, such as latent spaces. These models sacrifice transparency and interpretability for prediction accuracy. Such a DL algorithm is a Convolutional Neural Network (CNN). This is one of the most widely used models, forming the foundation for numerous consumer-facing applications. For example, in facial recognition products, CNNs use a learning process that transforms pixel-level inputs (an image) through complex connections across layers of the network, yielding highly arbitrary associations between inputs and outputs. This results in an inherently uninterpretable model for human users [4] [5].

Other examples of uninterpretable DL models are Transformer models, which are often used for natural language processing and form the basis of tools like ChatGPT, and Diffusion models, which are the basis for tools like DALL-E2, Stable Diffusion, and Midjourney.

Uninterpretable models are also referred to as “black box models” [3]. The interaction with these models is viewed as opaque, and the inner reasoning is frequently unknown and obscure [6] [7]. This opaqueness can hinder users’ trust in the system and discourage collaboration and usage, especially in situations where the system is in charge of high-risk decisions that can have significant consequences. Even though these models are exceptionally accurate [8], having an understanding of the reasoning behind such a critical decision is crucial [6], but the systems lack explainability and interpretability [8]. Ultimately, this distrust can also cause complete rejection of these systems [5]. These limitations have raised the need for the development of models that offer more clarity and transparency in order to improve the potential for interactions with these systems [9]. Addressing this issue, the field of XAI proposes to make a shift towards turning, specifically these inherently uninterpretable (or black box) models, into more transparent and interpretable AI systems. Initiatives like the Explainable AI program by the Defense Advanced Research Projects Agency (DARPA) [10], promote the need for XAI to further understand, trust, and manage the emerging generation of artificially intelligent machines. Gerlings et al. [11] support the importance of the generation of trust, transparency, and understanding as one of the main drivers of XAI systems.

As discussed in Section 1, establishing a better understanding of the workings of a system is critical when it comes to AI tools that are targeted at the general consumer. Another factor that plays a role in this understanding, according to Gerlings et al. [11], is the minimization of biases and misinterpretations in model performance and interpretation. Similarly, when it comes to these consumer-focused AI applications, reducing misinterpretation of both the workings and the results of the system is crucial to prevent the adaptation and spread of false information or biases.

Examples of commonly used XAI techniques that aim to open up black box models are LIME (Local interpretable model-agnostic explanations) [12], SHAP (SHapley Additive exPlanations) [13] and LRP (Layer-wise Relevance Propagation) [14]. Although these methods are successful in their goals of providing their users with insights into how the model reached a certain point or came to a specific decision, the results are still complex to understand and not accessible to users outside of the academic field, let alone non-expert users such as general consumers of daily-use AI tools.

### 2.2 Explainable Computational Creativity (XCC)

Another area where XAI is being explored is the field of Computational Creativity (CC). In this subfield of AI, computational systems are developed and used to create artifacts and ideas that are applied in domains associated with creative people, such as mathematics, poetry, music composition, video games, graphic design, and the visual arts. Colton and Wiggins [15] define Computational Creativity as "The philosophy, science, and engineering of computational systems that, by taking on particular responsibilities, exhibit behaviors that unbiased observers would deem to be creative."

Generally, the field of CC focuses on computer programs that autonomously produce works of art, classified as creative, with no human in the loop. However, a popular application of CC is Human-Computer Co-Creativity: a collaboration between humans and computers improvising in real time to generate a creative product. Important here is that the computer does not follow a predefined script as a guide for the interaction but rather adapts to the input of the human and generates responses based on computationally creative algorithms. In Human-Computer Co-Creativity, creativity emerges through the interaction of both the human and the computer. [16]

Llano et al. [17] introduce the application of XAI to the field of CC as Explainable Computational Creativity (XCC). They argue that in order to have fruitful co-creation with CC systems, there is a need for ongoing interaction and two-way communication. To facilitate this, the CC system has to be designed in a way that it can communicate and explain its processes, decisions, and ideas throughout the creative process in a way that is understandable to humans. The form of communication could be in the broadest sense, not just through linguistic forms but it should be fitting for the creative task and the information that is being communicated.

The method of explainability as proposed by Llano et al. [17] differs much from the methods used in the field of XAI (Section 2.1). XCC is more focused on the interaction (or, in their case, collaboration) between the user and the system, in a way that is intuitive for humans. Even though the methods are targeted at different types of systems, with XCC mainly focusing on co-creative models, this new approach and the proposed principles could be a step away from the complexity of XAI models and towards a wider audience of non-expert users.

### 2.3 Interactive Installation Art (IIA)

In the realm of Interactive Installation Art (IIA), artists have been exploring and utilizing engagement with an audience of non-experts for a long time. Technology is often integrated as a tool to invite participants to interact with the installation. Generally, IIA works rely on participant interaction to unlock their full potential encouraging individuals with no prior knowledge of the subject to actively engage with the installation through the used technology. Two examples, David Rokeby's "The Giver of Names" [18] and Rafael Lozano-Hemmer's "Pulse Room," [19] are explored to illustrate the experiences offered by IIA and how they invite viewers to become active participants. Subsequently, relevant studies and theories from the fields of Interactive Art and Installation Art will be introduced.

First, "The Giver of Names" [18] is an installation created by David Rokeby (Figure 2), who is known for his exploration of interactive and generative systems. In this installation, Rokeby utilizes computer vision technology to analyze and interpret the visual world. The artwork invites participants to arrange one or more objects on the pedestal, serving as the input for the installation. The computer program examines this live video input and attempts to assign names to the objects and compositions it detects. Through this process, Rokeby explores the relationship between language, perception, and technology, questioning how humans assign meaning and make sense of the world around us. The installation encourages participants to contemplate the limitations and possibilities of machine perception and the role of human interpretation in understanding our surroundings.

Another example is Rafael Lozano-Hemmer's "Pulse Room" [19] (Figure 3), an interactive installation with hundreds of clear incandescent light bulbs hanging from the ceiling in a dark room. The bulbs respond to participants' heartbeats detected through a sensor-equipped interface positioned at the opposite side of the room. It is only when a participant initiates the interaction by holding the interface that the meaning of the installation unfolds. The flashing of the light bulbs, previously meaningless, now starts to sync with the individual's heartbeat. This interaction reveals the true nature of the heartbeat sequence of the bulbs by briefly turning off all lights before the flashing sequence advances to the next bulb. Each interaction records the participant's heart pattern, which is transmitted to the first bulb, pushing existing recordings forward. This creates a real-time display of the collective recordings of previous participants. Essentially, installations like David Rokeby's "The Giver of Names" and Rafael Lozano-Hemmer's "Pulse Room" utilize technology in a way that encourages user participation, inviting viewers to contribute to the artwork and become part of the creative process.

**Fig. 2.** The Giver of Names [18]

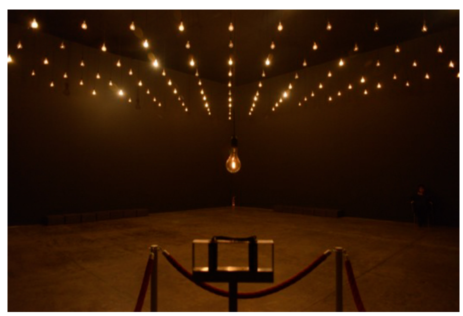

**Fig. 3.** Pulse Room [19]

As these examples demonstrate, IIA does not only foster user engagement but also provides an accessible gateway for people to interact with technology. By integrating technology within the artwork, these installations create opportunities for users to directly interact with and explore technological elements in a non-intimidating environment. This direct encounter with technology through interactive art installations removes intricacies that may exist when approaching complex technological systems. It offers an inclusive and user-friendly environment, regardless of a user's technical background, to actively engage with and develop a deeper understanding of the properties of the technology.

To better understand the properties behind the engaging nature of IIA, related studies and theories from the fields of Installation Art, Interactive Art, and Human-Computer Interaction are explored. First, Installation Art, as described by Bishop [20], encompasses artworks that the viewer physically enters, treating the space and its elements as a unified entity. It differs from traditional art forms by directly addressing the viewer's physical presence and multiple senses. Reiss [21] emphasizes the importance of spectator participation in Installation Art, which can take various forms, including interaction with specific activities or simply walking through the space and confronting what is presented. The recipient's interaction and engagement with the installation contribute to the evolving meaning of the artwork.

The concept of spectator participation extends beyond the art space, as demonstrated by Mast's [22] exploration of Augmented Play Spaces (APS). The Participation Journey Map (PJM) developed by Mast examines the motivations and influencing factors for engaging in APS, which integrates interactive technology into existing environments to promote playful interactions. The PJM identifies different states of engagement, such as Transit, Awareness, Interest, Intention, Participation, and Finishing, organized into the Onboarding and Participation phases. This shift from passive spectatorship to active user participation is further explored by Huhtamo [23], highlighting this as the most essential characteristic of Interactive Art. The viewer becomes an "interactor" who activates and realizes the artwork's full potential through physical and intentional engagement. Additionally, Kluszczynski [24] proposes four more defined roles that this interactor can assume: participant, performer, executor, and co-creator. Recipients can fulfill one or multiple roles ranging from experiencing and exploring the artwork to actively shaping its outcome and collaborating with the creator(s).

To describe the initial configuration of possible actions that are expected of the participant, as intended by the creator of the artwork, Kwastek [25] introduces the concept of the "Interaction Proposition". Furthermore, Kwastek explores the aesthetic distance in Interactive Art, noting that this field of art challenges the traditional static condition of aesthetic distance. Rather, in Interactive Art, the aesthetic distance fluctuates between immersion in the experience and a detached (self-)perception. This fosters the participant's self-awareness and evokes conscious reflection. Finally, Kwastek identifies four modes of experience in interactive artworks: Experimental Exploration, Expressive Creation, Constructive Comprehension, and Communication. In Section 4, these theories will be further explored and integrated into the phases of the framework. Ultimately, these theories contribute to a deeper understanding of the components and characteristics that shape Interactive Installation Art.

This section has provided a brief introduction to XAI, XCC, and IIA. In the next section, the intersection between the goals of XAI, the more human-centered and co-creative approach as proposed in XCC, and the potential of increasing non-expert engagement from IIA, will be further explored in the form of a preliminary study. Additionally, the preliminary study aims to gain insights into the current attitudes and understanding of AI systems of the target group.

## 3 Preliminary Study

This section presents the findings of a preliminary study conducted to gain initial insights into current attitudes and understanding of AI systems and to evaluate the potential use of IIA as a method for increasing engagement with non-expert users. The quantitative study involved an online survey with 84 participants recruited through social media platforms. The study aimed to gather information on participants' experience and openness to adopting new technologies, preferred methods of familiarizing themselves with new tools, and their preferences for aesthetic experiences. The findings suggest that participants had limited experience and understanding of AI image generation systems, but expressed openness to adopting new technologies. Furthermore, the preference for interactive experiences, especially in the physical realm, and the incorporation of displaying intermediate results of the image generation process were highlighted. Overall, these results provide valuable insights into the target audience's perspectives and preferences, informing the development of the framework for the utilization of IIA for XAI.

### 3.1 Study Design

The study had three primary goals. The first goal was to gather information regarding the average experience with, understanding of, and confidence in using AI image generation tools across different age groups and levels of tech-savviness. An additional objective was to gain insights into the openness towards adaptation of new tools and technologies. This information would help to better understand the target audience and tailor the final system to their backgrounds and attitudes.

The second goal was to explore the effectiveness of interactive experiences in physical space as a means of familiarization. Specifically, the study aimed to investigate the preferred methods of familiarizing oneself with a new tool or technology as well as the preferred mediums for evoking an aesthetic experience. This would help to further explore the potential utility of IIA for XAI. The final goal of the survey focused on gathering preferences related to the explainability of the AI image generation process. The primary question addressed the preferred approach for visualizing the process of image generation. These preferences could be further used in the final system, as a visual method of adding explainability to the model.

Participants were recruited via social media platforms such as LinkedIn and Instagram and other digital channels. The survey consisted of 15 questions, divided into five parts. The first part served to collect some general background information on the participant, like age, field of work/study, and tech-savviness.

The second part contained questions about the participant's current experience with AI image generators and their understanding and confidence in the usage of these tools. The third part contained three questions about the participant's preferred medium of experiencing art, their preferred medium for familiarization with a new tool or technology, and a rating regarding their general openness to adopting new tools or technologies. Part four showed the participant two examples of visualizations of the AI image generation process, after which the participant had to express their preference out of the two when it came to showing the progress of generation, showing how the image is being generated, and a personal preference for the overall attractiveness. Finally, the last part contained some open questions, giving the participant the opportunity to express their general attitude towards AI art and image generation and leave any other comments.

### 3.2 Results

In total, the survey was filled out by 84 respondents. The average age of the respondents was 39.7 and the average tech-savviness was 6.0. For the processing of the results, the respondents were divided into two age groups: group one with respondents between 16 and 40 (47 participants), and group two with respondents between 40 and 76 (37 participants). Participants were also grouped on tech-savviness, by dividing them into two groups based on their own rating on a scale of 0 to 10. This resulted in a group with 51 tech-savvy respondents, with scores higher than five, and a not-tech-savvy group with 33 participants, with scores of five or lower.

The average experience with AI image generators of all participants was quite low (2.4 out of 10), and 45% of the participants indicated that they had never used an AI image generator before. 20% said they had used DALL-E 2 before, and the rest had used generators like Craiyon, Stable Diffusion, or other tools. The average experience score was a bit higher in the younger age group (3.1), with 34% of inexperienced participants, than for the older participants (1.5), where 62% never used an AI image generator. Furthermore, tech-savvy participants also scored higher (3.2) than non-tech-savvy participants (1.0), out of which 68% of the latter group never used an AI image generator. In all groups, DALL-E 2 was the most commonly used AI image generator, followed by Craiyon in the not-so-tech-savvy and younger age groups and by Stable Diffusion and Midjourney in the tech-savvy group.

Participants rated their understanding of these systems in a similar distribution, but in between the other two with a total average of 3. Participants rated their confidence in using these tools a bit higher with an average of 3.5 in total (Table 1). Notable is that even the group with tech-savvy participants showed relatively low confidence in the use of these systems (4.5). Finally, participants gave remarkably higher scores for their openness toward using new tools and technologies, with an average of 6.3. This also goes for the older age group, which has mostly no experience with AI image generators but still scores their openness to new tools and technologies with a 6.0. Interestingly, of these three factors (experience, use-confidence, and understanding), all groups of participants scored experience lowest, followed by understanding, and gave the highest score to their confidence in using the system.

| | Age | Tech-savviness | Experience | Confidence | Understanding | Openness | N |
|---|---|---|---|---|---|---|---|
| **Age =< 40** | 27,1 | 6,4 | 3,1 | 4,4 | 3,9 | 6,4 | 47 |
| **Age > 40** | 55,7 | 5,6 | 1,5 | 2,4 | 1,9 | 6,0 | 37 |
| **Tech-savvy** | 37,9 | 7,6 | 3,2 | 4,5 | 3,8 | 7,0 | 51 |
| **Non tech-savvy** | 42,4 | 3,7 | 1,0 | 2,1 | 1,9 | 5,2 | 33 |
| **Total** | 39,7 | 6,0 | 2,4 | 3,5 | 3,0 | 6,3 | 84 |

**Table 1.** Respondent's average age and tech-savviness, experience with and confidence in using AI image generation tools, and openness towards adopting new tools

To explore the presumption that the use of IIA could be a fitting method to create engagement, the participants were asked to pick their preferred medium(s) of experiencing art and their preferred method of familiarizing themselves with a new technology or tool. The results are presented in Figure 4. Physical works scored much higher than digital works, with a total of 79% of votes from all participants and a preference of 91% in the non-tech-savvy group. Within the realm of physical works, dynamic and static works ranked similarly, while interactive works scored a bit lower in most groups except for the group with younger participants. Digital works were most preferred in the tech-savvy group (27%), specifically interactive (11%) and dynamic (9%) works. Half of the total participants chose interacting with the technology or tool as their preferred way of familiarizing themselves. Second to that, the most popular method was to watch a video about the technology/tool (30%). Asking a friend for an explanation got 12% of the votes, and reading a manual about the technology or tool was the least preferred, with only 8%. The latter was much more popular within the younger and tech-savvy age groups (13% and 14%) compared to the older and non-tech-savvy groups (3% and 0%). These results illustrate the case for a work in the physical space, either dynamic, static, or interactive, that allows users to interact with the tool/technology and/or incorporates some kind of video formatting to help users familiarize themselves with the technology.

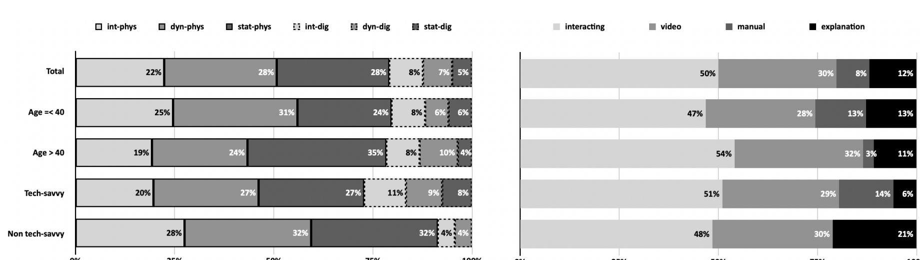


**Fig. 4.** Distributions for preferred medium for aesthetic experience (left) and preferred method for familiarization with new tools (right)

Finally, the preferred approach for visualizing the process of image generation was addressed. As transparency is shown to be a common theme in XAI (as explored in Section 2.1, the hypothesis was that showing the intermediate results of the image generation process would be a more fitting method of helping the user develop an intuition for the workings of the system. To test this hypothesis, two videos of two progress visualizations were presented to the participants: (1) a progress bar as used in most systems, in this case taken from DiffusionBee[1], and (2) a progress visualization showing the noise diffusion process (Figure 5). The participants were asked to indicate which example (one or two) they found to do a better job in terms of visualizing the progress, showing the generation process, and which one they preferred to look at while waiting for the final image.

The results were as follows: 63% preferred the second visualization for showing the progress while the image is being generated; 89% preferred the second visualization for showing how the image is being generated, and 73% preferred the second visualization to look at while waiting on the final image. These percentages indicate a clear preference for the second visualization showing the intermediate results of the generation process, particularly in terms of showing how the image is being generated. This suggests that incorporating this type of visualization in the installation could indeed contribute to an increased understanding and approachability of the technology. There also seems to be a slight advantage to this example in terms of showing the progress while the image is being generated, but it is not as noticeable. Also, almost three-quarters of participants preferred this example to look at while waiting on the final image.

**Fig. 5.** Some steps of the progress visualization showing the noise diffusion process

In conclusion, the preliminary study shows an overall low experience with (2.4/10), understanding of (3.0/10), and confidence in the usage of (3.5/10) AI image generation systems, especially in the age group above 40. Respectively, the results in this age group are 1.5/10, 1.9/10, and 2.4/10. In contrast, both age groups appear to be quite open to adopting new technologies. The average score for openness for the age group above 40 is 6.0/10 and for the group under 40, it is 6.4/10. This supports the notion that this group of (potential) users is generally open and willing to adopt and learn about new technologies but is not yet given the right tools and opportunities to engage with these systems. Additionally, the preference for interaction as a method of familiarization with a new tool was

[1] https://diffusionbee.com/

confirmed by 50% of participants, followed by watching a video about the tool, which was chosen by 30%. Moreover, 79% of participants selected the physical space as their preferred environment for aesthetic experience. Finally, it supports the notion that showing the intermediate results of the image diffusion process indeed provides a preferred method (89%) of presenting the generation process.

## 4 IIA4XAI Framework

This section introduces the Interactive Installation Art for Explainable Artificial Intelligence (IIA4XAI) framework. The framework provides a structured approach to understanding the dimensions of IIA and the dynamic relationship between the recipient and the artwork. The application of the framework to XAI aims to enhance users' understanding and interaction with AI systems by leveraging the principles of IIA. The IIA4XAI framework outlines five distinct phases covering the full lifecycle of the experience: Encounter, Observation, Interaction, Explanation, and Reflection. When applied to the scope of IIA (Figure 6), the Observation and Explanation phases are optional, as the Observation phase is dependent on the presence of other participants, and the Explanation phase is not applicable to art installations that do not intend to explain their contents.

However, when applying the framework to XAI (Figure 7), the installation inherently intends to be explainable. This makes the Explanation phase required and leaves only the Observation phase as an optional phase. The framework focuses on key aspects that promote user engagement and comprehension in interactive art installations and potentially XAI systems. It emphasizes the user's active role in interacting with and exploring the AI system, which corresponds to the role of a participant in interactive art installations. Applying the principles of IIA through this framework, allows the creation of user-centered and engaging experiences with XAI systems to promote transparency, trust, and comprehension, enabling users to explore and understand AI technologies better. Following, the different aspects of the framework will be further explored.

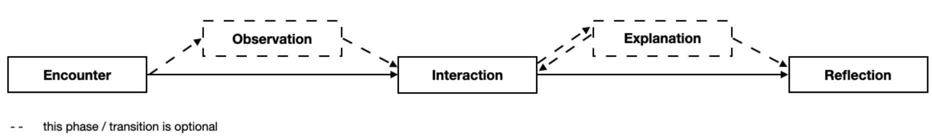


**Fig. 6.** Phases of user experience for IIA

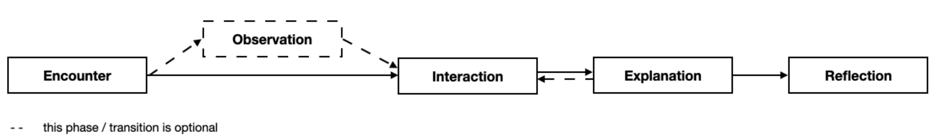


**Fig. 7.** Phases of user experience for IIA4XAI

### 4.1 Phases

The first phase of the framework is always the Encounter phase. This phase marks the recipient's initial awareness of the artwork, capturing their attention and curiosity. During this phase, it is important that the installation has a notable appearance, allowing the recipient to become interested in approaching the artwork. When there are already other recipients interacting with the installation, the recipient has the chance to transition to the Observation phase, where they switch their role from being a recipient to that of an observer. In this phase, the recipient (or observer) gets to take in the overall environment and setup of the installation. Additionally, the observer catches an initial glimpse of the possible interactions and experience with the installation. After the optional Observation phase, or directly after the Encounter phase, when the recipient (or observer) decides to actively approach and engage with the artwork, the Interaction phase is entered. During this phase, the recipient starts manipulating and exploring the installation's interactive features and takes on the role of a participant. The participant remains in this phase while the interaction is still taking place.

When the participant finishes interacting, they enter the next stage. Within the scope of IIA Figure 6 this next phase could be either the Explanation or the Reflection phase. In the Explanation phase, additional information or context about the technology or workings of the installation may be provided, to deepen the participant's understanding of the artwork, the interaction, and the technology. When it comes to IIA, not every art installation shares the goal of explanation, making this phase optional. When using IIA as a technique for XAI, however, the Explanation phase is critical for user engagement and fostering understanding of the technology, and is made compulsory (Figure 7). Furthermore, it is essential that the Explanation phase always takes place after the first cycle of the Interaction phase, as this allows the participant to initially develop their own unbiased interpretation of the artwork. The medium and formatting of the explanation in this phase are completely dependent on the technology in question and should be shaped to fit the properties of the model. Finally, the participant reaches the Reflection phase, where introspection and further processing take place. Participants are encouraged to reflect on their experience with the work and derive meaning and understanding from the installation.

The path pictured in Figure 7 is the main path. However, recipients are free to move back and forth between phases and are specifically encouraged to re-enter the Interaction phase after the Explanation phase in order to fully experience the installation in additional modes of experimental exploration and even expressive creation (Section 4.2). This is specifically the case in scenarios where there are multiple participants present at the same time which could lead to collaborative interaction. This collaboration specifically enhances the Reflection phase, as it encourages conversation and discussion between participants, comparing experiences and interpretations regarding the work.

### 4.2 Components and Characteristics

In this section, the framework is expanded with relevant theories and studies from the fields of Installation Art and Interactive Art, as introduced in Section 2.3. First, in Figure 8 the current theories are explored and applied to the framework by translating them into components and laying them out over the Phases that they apply to. Subsequently, in Figure 9 the framework is expanded by the introduction of possible factors or characteristics as explored in this section. These factors could be used as building blocks for the installation and can be tied back to the five phases. These theories and their characteristics will be further elaborated upon in divided subsections.

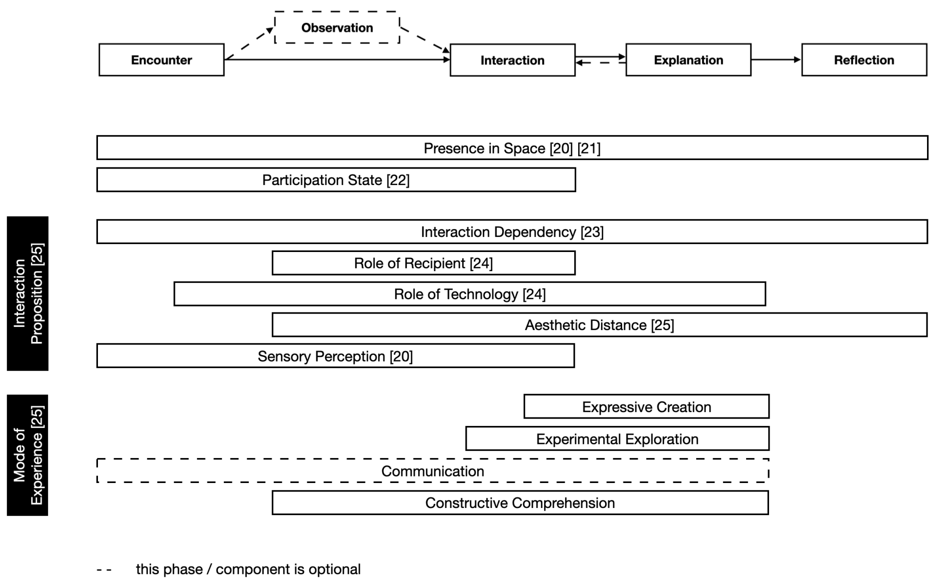


**Fig. 8.** Components for IIA, distributed over the phases

**Presence in Space** Installation art encompasses all art into which the viewer physically enters and is often referred to as 'theatrical', 'immersive', or 'experiential' [20]. What distinguishes installation art from other art is that the space and the ensemble of elements within it are considered in their entirety as a singular entity. In contrast to traditional art forms (such as sculptures, paintings, and videos), installation art addresses the viewer directly as an actual presence in the space and, with that, assumes an embodied viewer, not only with an aware sense of vision but also those of touch, smell, and sound. [20]

Reiss [21] also names this direct presence of the viewer, or what she calls: spectator participation, as the essence of installation art. She argues that the definition of participation varies greatly among artists; participation can mean

| | | Encounter | Observation | Interaction | Explanation | Reflection |
|---|---|---|---|---|---|---|
| | Presence in Space [20] [21] | passing by [22] entering [20] | spectating [21] | confronting / exploring / spectator participation [21] | the recipients' presence in the space is an important factor in these phases, but can takes on a fluid form | |
| | Participation State [22] | transit / awareness / interest / intention | interest / intention | exploration / continuation | exploration / continuation / finishing | finishing |
| interaction Proposition [25] | Interaction Dependency [23] | activation | activation | activation | fulfillment [21] | fulfillment [21] |
| | Role of Recipient [24] | | observer | participant / performer / executor / co-creator | observer | |
| | Role of Technology [24] | performer | performer | participant / performer / executor / co-creator | performer | |
| | Aesthetic Distance [25] | distanced perception | distanced perception | immersion / distanced (self-)perception | detached evaluation / distanced (self-)perception | detached evaluation / distanced (self-)perception |
| | Sensory Perception [20] | visual / auditory / haptic / olfactory | visual / auditory / haptic / olfactory | visual / auditory / haptic / olfactory | visual / auditory | |
| | Mode of Experience [25] | communication | communication / constructive comprehension | communication / constructive comprehension / experimental exploration / expressive creation | communication / constructive comprehension / experimental exploration / expressive creation | constructive comprehension |

**Fig. 9.** Components and their characteristics, per phase

offering the viewer specific activities, but it can also mean that the viewer is simply demanded to walk through the space and confront what is there. In most cases, participation takes the form of interaction, thereby categorizing the artwork as an interactive artwork. Furthermore, the way in which objects are presented to the viewer can also differ; objects may fall directly in the viewer's path, but an exploration of the space can also be needed in order to uncover them. However, in all cases, the viewer is required to fulfill the artwork; "The meaning evolves from the interaction between the two".

**Participation State** A good example of how this spectator participation is being used beyond the art space is that of Augmented Play Spaces (APS). APS are installations in existing, (semi-)public environments that aim to promote playful interaction through the integration of interactive technology. In the Participation Journey Map (PJM), Mast [22] explores what motivates people to engage in these Augmented Play Spaces and what (design) factors can have an influence on people's decisions to engage or participate. The PJM presents six (potential) states of engagement that participants experience while coming into contact with an APS: Transit, Awareness, Interest, Intention, Participation, and Finishing, grouped by two main phases: Onboarding and Participation. Figure 9 shows how these states of engagement translate to the phases of the IIA4XAI framework.

The first state, Transit, refers to the state of a potential participant as a passer-by. From this state on, when the passer-by encounters the APS (or installation), they become knowledgeable of its existence and therefore enter the next state: Awareness. A key factor in this state is the noticeability of a system. Once a potential participant becomes curious, they enter the Interest state. Reaching this state can be strengthened by enabling them to observe the system, preferably with active participants. When the potential participant's interest in the installation reaches a decision to participate, the next state of Intention (to participate) is entered. This is where the Onboarding phase ends, and the passer-by becomes a participant.

The Participation phase starts with the state of Exploration, facilitated by curiosity, usability, and reward. To encourage participation, it is, among other factors, important that the installation is inviting, has a low entry barrier, is intuitive to use, and has layered content. From this point on, when the participant has explored the interaction with the system, they have two options: continuing, or finishing the interaction. Encouraging Continuation can be done by supporting different roles and interactions, competition, giving participants autonomy, facilitating emergent play, and offering new challenges, experiences, and multiple layers of complexity. Finally, when a participant decides to end the interaction, the Finishing state is entered. The intention to stop the interaction can be caused by both intrinsic motivators (like feeling finished, boredom, or a lack of progress) and extrinsic motivators (like running out of time or having reached the end of the interaction).

**Interaction Proposition** Where in traditional art the work is the primary object and the viewer is passive, interaction art requires the visitor to take on an active role as a user or 'interactor' [23]. Interactive Art is based on an 'artistically configured interaction proposition' [25], most often appears in the form of an installation [23] and requires the user to put in physical effort to fulfill the artwork. In Interactive Art, the artwork depends on the user's action to activate, function, and realize its full potential. This will be further referred to as the "Interaction Dependency" in Figure 8 and 9. This shift of the visitor from the role of just a spectator to the role of user or 'interactor' or vital factor in the Interaction Proposition, aims to "empower and challenge the visitor to go beyond the modes of usual [passive] spectatorship" [23] and is essential for interactive works. The Interaction Proposition encompasses all components and characteristics of the artwork that are composed by the artist, with the intention of provoking the participant's interaction. The components of "Role of Recipient and Technology", "Aesthetic Distance", and "Sensory Perception" as described in the following sections, are all part of the Interaction Proposition.

**Role of Recipient and Technology** Kluszczynski [24] suggests that a recipient of an interactive work can take on one or more of the following roles: participant, performer, executor, and/or co-creator. In the role of participant, the recipient engages with the work by experiencing and exploring its features and functionality. The participant does not have an active role in terms of shaping or creating the work but rather interacts with what is presented. In the role of performer, the recipient actively engages with the work by manipulating and creating content within it. The performer's actions play an active role in determining the outcome and other recipients' experience of the work. In the role of executor, the recipient is tasked with executing specific actions within the work. These actions, like activating or controlling certain elements of the work, may be predetermined by the artist, or assigned by other participants or (co-)creators. Finally, in the role of (co)creator, the recipient actively collaborates with the creator(s) of the work in order to develop, shape, or refine its features, functions, and overall aesthetic experience. The (co-)creator actively contributes their ideas and creative input, becoming an essential part of the creative process.

The four roles described above are not mutually exclusive, meaning that a recipient can take on different or multiple roles at the same time or over the course of their engagement with the piece. When applied to an interactive work that incorporates technology, the technology is considered as another actor. Therefore, the technology can also take on any of the roles mentioned. As can be seen in Figure 9, the role that both the recipient and the technology can take on depends per phase.

**Aesthetic Distance** In art, the concept of aesthetic distance is used to refer to the psychological and emotional space that exists between the recipient and the artwork. Aesthetic distance represents the degree of detachment that provides a mental and emotional buffer, allowing recipients to observe the artwork's elements and interpret its meaning objectively. In traditional visual arts, such as paintings or sculptures, the aesthetic distance is associated with the physical distance between the position of the viewer and the artwork. This results in the aesthetic distance as a static condition.

On the contrary, in interactive art, this static condition of aesthetic experience does not exist. As the recipient actively engages with the artwork, the boundaries between the recipient and the artwork become blurred. Rather, the aesthetic distance appears as a fluctuation between the establishment of artificiality and reality, and between a state of immersion, in the activity and distanced (self-)perception [25]. By building on this tension between the sense of being fully immersed in the experience and stepping back to observe and evaluate the experience from a more detached perspective, a form of self-awareness comes into practice where the recipient is simultaneously participating in and observing their own interaction with the work. These fluctuations of aesthetic distance are laid out over the phases in Figure 9.

**Sensory Perception** By inviting the participant to physically interact with the artwork, new sensory dimensions are introduced to the experience. As the physical interaction commonly invites the user to physically touch the artwork, the "haptic dimension" [23] is added to the frequently already present visual dimension. The auditory dimension is also commonly introduced by the artwork's use of sound. Finally, the olfactory dimension can be added by the usage of smells. The use of multiple senses allows for a more immersive and engaging experience for the recipient, helping them to go beyond the modes of traditional spectatorship [23]. As can be seen in Figure 9, the haptic and olfactory dimensions are usually only incorporated in the first three phases (Encounter, Observation, and Interaction) as they are not as suited for conveying information in the Explanation phase, and do not add to the recipient's personal reflection in the Reflection phase.

Understanding the work requires more than just sensory perception; it also requires an active cognitive process in the form of either emotional or physical transformation or conscious reflection. Kwastek [25] suggests that the aesthetic experience of interactive art is created through a combination of instrumental arrangements, physical and visual compositions, activated by the recipient(s), and their placement within the recipient's personal experiences and frame of reference. The interpretation of the aesthetic experience varies from recipient to recipient based on subjective perception and the construction of meaning. Ultimately, intellectual understanding of the art is always based on an interplay between sensory perception, purposeful action, cognitive knowledge, and the recipient's active individual experience.

**Modes of Experience** According to Kwastek [25], a system can be experienced through four different Modes of Experience, the first being Experimental Exploration. When recipients engage with an interactive system, they seek to explore its fundamental rules and properties. Various methods can be used to increase awareness during this process of experimental exploration. One such way is through disruption, which involves strategies that challenge the recipients' control of the system and prompt them to explore and learn more. According to Feingold [26], a recipient often wants to understand how the system works, if they are using it correctly, and what they have to do in order to achieve a specific outcome. By disrupting the expected experience, epistemic processes are evoked. Another strategy is repetition, where recipients are encouraged to repeat actions and create distance from their own actions by considering them as just one of many possible behaviors. This behavior can be built into the structure of the system itself, or it can be explicitly promoted.

The second Mode of Experience is that of Expressive Creation. After a period of experimental exploration, once recipients become more comfortable with the mechanics of the system, they can be encouraged to generate something new from it. This leads recipients to consciously engage in their own creative process, and by gaining a deeper understanding of the system's rules and principles, they can deliberately apply these rules for further expressive creation.

The third Mode of Experience, Constructive Comprehension, not only includes discovering and creatively utilizing the interaction possibilities, but also exploring and configuring the symbolic aspects of the system. Constructive Comprehension becomes specifically relevant when an interactive system involves some type of spatial configuration, such as physically or virtually moving through space, where the space is critical for the contextualization of the system.

The last Mode of Experience is Communication which entails any feedback processes that recognize recipients as rational individuals and engage them in a mutual exchange of ideas. It is important to emphasize that this definition includes not only real-time exchanges but also other forms of communication, such as observation. Whereas the other Modes of Experience are usually present in at least one of the phases of the framework, the mode of Communication does not always exist. This is also illustrated in Figure 8.

Ultimately, the framework presented in this section outlines five essential phases for IIA: Encounter, Observation, Interaction, Explanation, and Reflection. By adapting this framework to the domain of explainable AI systems, engaging user experiences can be created that enhance understanding and interaction with AI technologies. The framework emphasizes the user's active role, involvement of all senses, cognitive reflection, aesthetic distance, and different modes of experience. Integrating elements of IIA into explainable AI design helps foster meaningful and immersive experiences, paving the way for a more accessible and user-centered approach to AI.

## 5 Case Study

To further explore and evaluate the IIA4XAI framework, a case study was done. The case study builds upon the preliminary study that was presented in Section 3. In the preliminary study the target group's tech-savviness, general experience with, understanding of, and confidence in the usage of AI image generation systems, and their openness towards adopting new technologies were investigated. Furthermore, preferred methods of familiarizing with new tools and environments for aesthetic experience were evaluated. Finally, the preliminary study compared two visualization methods for showing the image generation process.

The goal of the case study was to assess the effectiveness of the IIA4XAI framework as a foundation for the design and development of interactive art installations that serve as a first step toward XAI. A prototype for an interactive art installation was designed by using the phases, components, and characteristics of the framework as building blocks. The prototype was then used to conduct qualitative user tests on a small group of nine participants. The aim of this test was to evaluate the effect of the installation on the participant's understanding of, interest in, and engagement with AI image generation technology.

To gain insights into the initial attitude and background knowledge of the participants, the case study once more investigated the participant's tech-savviness, general experience with, understanding of, and confidence in the usage of AI image generation systems, and their openness towards adopting new technologies. Additionally, participants' initial interest in learning more about the workings of the system, and their feeling of engagement with the technology were examined. To assess the influence of the interactive installation, the case study further examines participants' understanding, interest, and engagement before and after interacting with the prototype. For evaluation of the general user experience, with a particular focus on the installation itself, the study incorporates the User Experience Questionnaire (UEQ+). Ultimately, the case study aims to provide valuable insights into the impact and effectiveness of the interactive installation and its potential for promoting engagement with the technology.

### 5.1 The Installation

The installation consisted of seven stacked old CRT television sets, with six facing the front (Figure 10) and one facing the back (Figure 11). The television set at the back was connected to a camera that pointed at a blank white wall, displaying the camera's input. Participants were invited to position themselves between the white wall and the camera, generating the input image that triggered the image generation process. The use of the participant's own face as input for the system allows for a personalized experience and helps to build engagement between the participant and the installation.

The camera input was processed using an adapted version of the CLIP Interrogator 2.1 model [2], which generated a prompt based on the participant's image. This prompt was then used as input for a Stable Diffusion-based pipeline [3], resulting in the generation of five completely unique images. Each television displayed a different generated image, looping through the various stages of the generation process, from the initial noise to the final result. This way of displaying the generation process came out of the preliminary study (as described in Section 3) as the preferred visualization. As participants moved closer, they could observe the evolution of each image, gaining insight into the AI generation process.

Once all five images were generated and displayed on the front televisions, the prompt that had been used for generation appeared on the larger television set positioned at the bottom of the installation. This prompt served as a textual representation, or explanation, of the participant's image and provided insights into the link between their own image and the generated results. In Appendix A (Figure A3 and Figure A4) examples of such an input image, the generated prompt, and the resulting images are shown. Additionally, Appendix A contains a schematic overview of the layout of the phases in the space (Figure A1) and the configuration of the installation (Figure A2). The full code can be found at https://github.com/vannoordenne/iia4xai.

**Fig. 10.** The front side of the installation

**Fig. 11.** The backside of the installation

The old CRT television sets were chosen as the medium for the installation because they induce a striking contrast between old analog technology and the modern state-of-the-art AI model. Furthermore, as also became clear from the preliminary study (Section 3), a majority of the target group covers users above the age of 40, as this age group showed the least experience with, and confidence in usage of, these types of AI systems. For this group of users, these old televi-

[2] https://github.com/pharmapsychotic/clip-interrogator
[3] https://huggingface.co/stabilityai/stable-diffusion-2

sion sets are generally a familiar medium from the past. Additionally, the noise that is displayed at the start of each generation process bears resemblance to the static noise displayed on CRT television sets with no signal, adding another link between the two technologies. This induces a sense of nostalgia, a technique that is commonly adapted in the field of marketing to evoke a sense of connection, a positive mood, and openness in the participant [27]. Hence, the use of this nostalgic medium serves to strengthen the connection between the participant and the installation, fostering engagement and motivation for further exploration.

For the overall composition of the participant's experience and interaction, the characteristics as illustrated in Figure 9 (Section 4 were used). Figure 12 shows the IIA4XAI framework together with the components and characteristics, as applied in the design of the case study installation. The components or characteristics in grey were not used in the installation.

| | | Encounter | Observation | Interaction | Explanation | Reflection |
|---|---|---|---|---|---|---|
| | Presence in Space [20] [21] | passing by [22] entering [20] | spectating [21] | confronting / exploring / spectator participation [21] | the recipients' presence in the space is an important factor in these phases, but can takes on a fluid form | |
| | Participation State [22] | transit / awareness / interest / intention | interest / intention | exploration / continuation | exploration / continuation / finishing | finishing |
| interaction Proposition [25] | Interaction Dependency [23] | activation | activation | activation | fulfillment [21] | fulfillment [21] |
| | Role of Recipient [24] | | observer | participant / performer / executor / co-creator | observer | |
| | Role of Technology [24] | performer | performer | participant / performer / executor / co-creator | performer | |
| | Aesthetic Distance [25] | distanced perception | distanced perception | immersion / distanced (self-)perception | detached evaluation / distanced (self-)perception | detached evaluation / distanced (self-)perception |
| | Sensory Perception [20] | visual / auditory / haptic / olfactory | visual / auditory / haptic / olfactory | visual / auditory / haptic / olfactory | visual / auditory | |
| | Mode of Experience [25] | communication | communication / constructive comprehension | communication / constructive comprehension / experimental exploration / expressive creation | communication / constructive comprehension / experimental exploration / expressive creation | constructive comprehension |

**Fig. 12.** The IIA4XAI components and their characteristics, applied to the case study

**Encounter Phase** When potential recipients are passing by the installation (in transit), they become aware of the installation both by the notable visual presence of the big stack of old television sets and camera setup and by the more subtle auditory cues in the form of high-pitched and static sounds coming from the old CRT television sets. If their interest is piqued and they develop an intention to participate, their presence moves into the space, and they transition to a state of entering.

Initially, potential recipients do not fulfill a role in the interaction proposition, as they are not yet part of it. On the other hand, because the installation is already displaying the results of the previous interaction in an attempt to gain the potential recipient's attention, the technology is taking on the role of the performer. When it comes to sensory perception, both the visual and auditory senses are present (and will stay this way over the course of the other phases). As the potential recipient is not immersed in or participating in the work yet, the aesthetic distance is perceived as a distance. There is no interaction dependency when it comes to the activation of the work, as the installation was already activated by the participation of previous recipients. Finally, as for the mode of experience, since this installation does not use any form of communication, there is no mode of experience present in this phase.

**Observation Phase** After a recipient decides to move into the space and enter the installation, the state of participation is unchanged, as the recipient still has an interest in and the intention to interact with the installation. The recipient starts passively observing the front side of the installation (Figure 13), becoming a spectator, and taking on the role of an observer in the interaction proposition. The technology, however, still fulfills the role of the performer by displaying the results of previous interactions. The potential recipient is still not immersed in or participating in the work, leaving a distanced aesthetic distance. As for the activation interaction dependency, the installation was already activated by the participation of previous recipients. However, for meaning to evolve, the recipient must interact with the work themselves in the following phase. Finally, since the recipient is able to familiarize themselves with the overall configuration of the space and the interaction proposition of the work, the constructive comprehension mode of experience is initiated.

**Fig. 13.** Recipient in the role of observer

**Interaction Phase** When the recipient moves further into the space and starts to approach the installation, their presence fluctuates between close confrontation and exploration of the space. By moving around the installation, the participant gets to investigate and eventually partake in the interaction proposition. By engaging in the interaction of the installation, the participant's participation state moves from the state of exploration into a state of continuation. Looking closer at the interaction proposition during this phase, by positioning themselves on the backside of the installation to generate input (Figure 14 and 15), the recipient has now transferred into the role of participant, but can also still take on the role of (active) observer. Furthermore, by providing input for the computational process of the installation and thus contributing to its final output, the recipient also fulfills the role of co-creator of the system.

The same goes for the role of the technology: as the system is responsible for the actual generation process of the final result images, the technology also plays a co-creating role in the overall installation. Nevertheless, the technology still fulfills the performing role by displaying the resulting images on television screens. When it comes to interaction dependency, the participant's personal activation of the system allows for the fulfillment of the installation and for meaning to evolve. Moreover, an additional mode of experience, besides the still present mode of constructive comprehension, is now introduced: the experience of experimental exploration. After the participant completes the first interaction cycle, from input to output, the participant is encouraged to repeat the interaction and thereby start exploring different relations and properties that they might influence by altering their input. In the phase of Interaction, by taking on different roles within the installation, the aesthetic distance fluctuates between a state of immersion and a state of distanced (self-)perception.

**Fig. 14.** Participant generating input

**Fig. 15.** Successful input generation

**Explanation Phase** After the generation of the images is finished, and all results are displayed on the television screens, the prompt that was derived from the participant's input is presented (Figure 16). This marks the beginning of the Explanation phase, where the participant can start to make connections between the different components and start investigating the properties and workings of the system. The participant's presence in space is quite fluid in this phase, as they can freely get closer to the television screens to analyze the images or the prompt (Figure 17), but they are also allowed to move around the installation or return to the backside to start another interaction cycle by generating new input. By doing this, the participation state can take the form of exploring or continuing. Furthermore, the participant can also decide to end their participation, moving into the state of finishing.

While in the Explanation phase, the interaction proposition involves the role of the recipient as an observer and the role of technology as a performer. By taking a step back into the role of observer again, the aesthetic distance takes on the form of detached evaluation or distanced perception. Interaction dependency exists, as the Explanation phase can only reach its full potential when the participant is responsible for the generation of input. Finally, the mode of experience continues with constructive comprehension, as explanation adds to the total comprehension of the experience. Experimental exploration can be even more fruitful with more explanation and can encourage participants to go back to the interaction phase, enabling expressive creation.

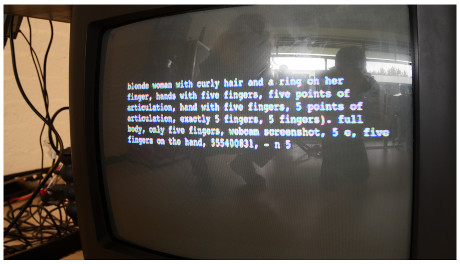


**Fig. 16.** The presented prompt

**Fig. 17.** Participant observing the prompt

**Reflection Phase** Finally, when the participant decides to enter the participation state of finishing the previous phase, the Reflection phase is entered. Presence in space is unimportant in this phase, as this phase mostly takes place in the form of cognitive processes. Additionally, most factors of the interaction proposition are not applicable in this phase, except for the aesthetic distance. As this phase is mainly about reflecting on and evaluating the experience and deriving meaning and understanding from it, the aesthetic distance is mainly that of detached evaluation and distanced (self-)perception. As for the mode of experience, only the mode of constructive comprehension is present at this stage, as the reflection adds to the overall comprehension of the experience.

### 5.2 Study Design

To gain insights into the influence of the installation on the user's engagement and understanding of the AI image generation technology, user tests were conducted. During the user tests, data was collected through observation and recordings of the interactions. Furthermore, participants were asked to fill out a questionnaire before, and after, interacting with the work.

The initial questionnaire consisted of seven general questions regarding age, educational/professional background, level of tech-savviness, experience level with regard to AI image generation systems, and openness towards new technologies. Additionally, to measure the starting point of each participant and the effects of the installation, participants were asked to rate themselves on a scale of 0 (none) to 10 (exceptional) in terms of understanding the workings of these systems, their interest in learning more about the technology, and their current feeling of engagement with the technology. After the interaction with the installation, participants were asked to fill out the second part of the questionnaire, in which they were again asked to once more rate themselves on a scale of 0 to 10 for the same three questions as before regarding their understanding, interest, and engagement with the technology. Furthermore, participants were asked to evaluate the overall usability and user experience of the installation by ranking twenty items spread over five scales based on the User Experience Questionnaire (UEQ+) [28]: Attractiveness, Stimulation, Novelty, Intuitive Use and Quality of Content. For example for the first item in the scale of Attractiveness the following question was asked: “In my opinion, the installation is generally ...” allowing the user to select one out of seven values (-3 to +3) on a scale from ‘annoying’ to ‘enjoyable’.

Nine participants were recruited, partly out of a group that showed interest in participating in further studies after the preliminary study, and partly via the researcher’s personal network. Participants were between the age of 21 and 60 years, with five participants younger, and four participants older than 40. The participants’ professions ranged from business consultancy and IT support to nursing and physiotherapy. None of the participants had any educational background or professional experience in the field of Artificial Intelligence.

### 5.3 Results

This section presents the results of the case study that was conducted to evaluate the effectiveness of the interactive art installation in enhancing users’ understanding, interest, and engagement with AI image generation technology. The results are divided into three sub-sections: questionnaire, user experience questionnaire (UEQ+), and observations. Overall, the results indicate that the interactive installation had a positive impact on users, improving their understanding, interest, and engagement with AI image generation technology.

**Questionnaire** In Table 2 a total overview of all results per participant of the qualitative part of the questionnaire is displayed. The first group of five columns includes the ratings of their own tech-savviness, previous experience with and confidence in the usage of AI image generation systems, and their general openness towards the adaptation of new technologies or tools. Additionally, in the three following groups, consisting of three columns each, the results of the participant’s understanding of the workings behind AI image generation, their interest in learning more about it, and their feeling of engagement with the technology, both before and after their interaction with the installation are depicted. For each of these three topics, ‘Understanding’, ‘Interest’, and ‘Engagement’, the difference between the ratings was calculated in the form of a ’Gain score’.

On average, participants showed a moderate level of tech-savviness (6.3) and had minimal prior experience (0.7). Their confidence in the usage of such systems was under average (3.8), however, their openness rated high (7.6). This illustrates that these non-expert participants are very open to adopting new technologies like this, but run into a major lack of experience and confidence in these systems. The understanding of the technology also rated low with an average of 2.8 and only a single participant with a mediocre understanding score of 6. After interacting with the installation, all participants demonstrated an increase in general understanding, with an average gain score of +3.3 points.

This resulted in an average understanding of 6.1 after the interaction and raised six (out of nine) participants to a mediocre understanding of 6 or higher. In percentage terms, the participants' general understanding increased by 108% on average, indicating a substantial improvement.

Additionally, all except one participant showed an increase in interest in learning about the workings of the technology. The average increase in interest of +1.6 or 31%, was not as high as the increase in understanding and engagement. Finally, most participants also indicated an increase in engagement of at least two points. It is noticeable that participants who started with an engagement score below 5 experienced the biggest increase in engagement after the experience. On average, the engagement increased by 79%. No apparent correlation between the gain in understanding, interest, and engagement was observed. Furthermore, it can be observed that all experienced an increase in at least one of the three areas.

Overall, these results indicate that the interactive installation had a positive impact on the connection between the participant and the technology, by more than doubling their understanding, creating more interest, and highly increasing the overall engagement.

| General | | | | | Understanding | | | Interest | | | Engagement | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Participant | Tech-savvy | Experience | Confidence | Openness | Pre | Post | Gain score | Pre | Post | Gain score | Pre | Post | Gain score |
| 1 | 7 | 5 | 8 | 7 | 4 | 8 | +4 | 9 | 10 | +1 | 3 | 8 | +5 |
| 2 | 6 | 1 | 0 | 6 | 0 | 2 | +2 | 6 | 7 | +1 | 1 | 7 | +6 |
| 3 | 8 | 0 | 5 | 9 | 4 | 8 | +4 | 8 | 9 | +1 | 3 | 10 | +7 |
| 4 | 7 | 0 | 7 | 9 | 3 | 6 | +3 | 7 | 7 | +0 | 6 | 6 | +0 |
| 5 | 7 | 0 | 0 | 6 | 3 | 5 | +2 | 4 | 8 | +4 | 3 | 5 | +2 |
| 6 | 7 | 0 | 7 | 10 | 6 | 7 | +1 | 7 | 9 | +2 | 7 | 7 | +0 |
| 7 | 7 | 0 | 5 | 7 | 5 | 6 | +1 | 6 | 8 | +2 | 5 | 5 | +0 |
| 8 | 4 | 0 | 2 | 5 | 0 | 5 | +5 | 3 | 5 | +2 | 0 | 6 | +6 |
| 9 | 4 | 0 | 0 | 9 | 0 | 8 | +8 | 8 | 9 | +1 | 6 | 9 | +3 |
| | | | | | | | | | | | | | |
| AVG | 6,3 | 0,7 | 3,8 | 7,6 | 2,8 | 6,1 | +3,3 | 6,4 | 8,0 | +1,6 | 3,8 | 7,0 | +3,2 |
| AVG (%) | | | | | | | 108% | | | 31% | | | 79% |

**Table 2.** Questionnaire results per participant, divided by categories: General, Understanding, Interest, and Engagement

**User Experience Questionnaire (UEQ+)** The overall results of the UEQ+ are plotted in Figure 18. It is important to notice that the y-axis of the plot actually has a minimum of -3. Since none of the scores included negative values, this lower half of the scale was left out for clarity. Additionally, a more detailed overview of the different items per scale, the means, and standard deviations, as provided by the authors of the UEQ [28] can be found in Table 3. As the number of participants was small, with a total of nine subjects, it is hard to derive significant conclusions. However, looking at the results can still provide an overall notion of how the installation was generally experienced.

In Figure 18 it can be observed that all scales generally scored high, with medians ranging between 1.75 and 2.75 (on a scale of -3 to 3). In most scales, the outliers are quite consistent. The scale of Novelty, however, does not show any outliers as all of the ratings fell into the range of 2.25 and 3. This means

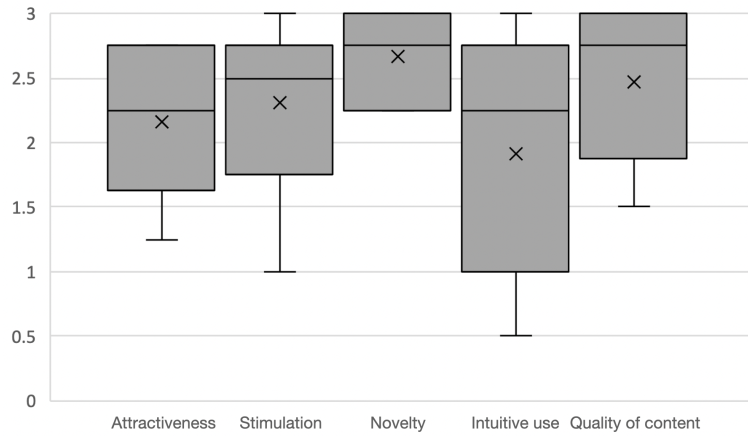


**Fig. 18.** Box-and-whisker plots per UEQ+ scale

that all participants agreed on a very novel experience, specifically in terms of creativity and innovation, and in being unconventional and leading edge.

On the other hand, some items falling into the scale of Intuitive Use, like ease of use and logicality, rated a bit lower. Ease of use specifically, had a very large standard deviation in comparison to the rest, indicating a big division in participant's perception of this topic. This could be caused by some complexities due to the system still being in a prototype stage, such as a long waiting time and an occasional jam in displaying the results.

| Scale | Item Left | Item Right | Mean | Variance | Std.dev. | N |
|---|---|---|---|---|---|---|
| **Attractiveness** | annoying | enjoyable | 2,44 | 0,78 | 0,83 | 9 |
| | bad | good | 2,44 | 0,28 | 0,50 | 9 |
| | unpleasant | pleasant | 1,56 | 1,03 | 0,96 | 9 |
| | unfriendly | friendly | 2,22 | 0,44 | 0,63 | 9 |
| **Stimulation** | not interesting | interesting | 2,78 | 0,19 | 0,42 | 9 |
| | boring | exiting | 2,22 | 1,19 | 1,03 | 9 |
| | inferior | valuable | 1,89 | 1,61 | 1,20 | 9 |
| | demotivating | motivating | 2,33 | 0,50 | 0,67 | 9 |
| **Novelty** | dull | creative | 2,78 | 0,44 | 0,63 | 9 |
| | conventional | inventing | 2,67 | 1,00 | 0,94 | 9 |
| | usual | leading edge | 2,56 | 0,53 | 0,68 | 9 |
| | conservative | innovative | 2,67 | 0,25 | 0,47 | 9 |
| **Intuitive Use** | difficult | easy | 1,56 | 3,78 | 1,83 | 9 |
| | illogical | logical | 1,56 | 1,53 | 1,17 | 9 |
| | not plausible | plausible | 2,22 | 1,19 | 1,03 | 9 |
| | inconclusive | conclusive | 2,33 | 0,75 | 0,82 | 9 |
| **Quality of Content** | obsolete | up-to-date | 2,56 | 0,53 | 0,68 | 9 |
| | not interesting | interesting | 2,89 | 0,11 | 0,31 | 9 |
| | poorly prepared | well prepared | 2,11 | 0,86 | 0,87 | 9 |
| | incomprehensible | comprehensible | 2,33 | 0,75 | 0,82 | 9 |

**Table 3.** UEQ+ mean scores per item

**Observations** When initiating the Interaction phase, the camera and the television screen displaying the camera's input on the backside of the installation, made it clear to all users that this was where the input generation would take place. They found it easy to understand the expectation of them taking place in front of the wall, facing the camera, and standing still until the display showed a sign of completion. However, participants two, four, and seven demonstrated confusion about how they were expected to pose in front of the camera. In these cases, they chose to position themselves in a very neutral and passive manner. Interestingly, these participants expressed the need for repeating the interaction, resulting in collaborative interaction.

After the input generation was done, and the first images started appearing, most participants started to analyze each image on an individual basis, looking for elements in the images that reminded them of themselves. For some participants, like one, four, five, six, and eight, this meant recognizing an overall image of how they would come across, like a professional or artistic appearance, or aesthetic features like hair color, ethnicity, or age. In contrast, participants three and seven could not find immediate similarities based on these impressions, so tried to look for them in more subtle elements, like their overall way of standing or other mannerisms.

On the other hand, some participants immediately recognized themselves in these more subtle elements, like their posture or way of smiling. This was even more interesting in the cases of participants two and nine, as they pointed out specific mannerisms, facial compositions, or hairstyles that were not captured on the input photograph. This suggests an interesting behavior, where participants expect some kind of background knowledge of the machine about themselves, but also puts a lot of focus on how they think that they come across.

After all images were generated, participants started to look for patterns across them. The focus was now mostly being put on specific colors, often coming from the colors of their clothing, compositions, and activities that appeared in multiple images. This became especially apparent in the interaction of participant three, who was wearing a red t-shirt and a black hooded jacket. These two colors came out to be extremely dominant in the final images, resulting in all images consisting mainly of black, red, and white colors, and all belonging to a very specific style. Subsequently, the participant concluded that "the database on which the model was trained probably contains a very strong relation between these colors and this style of image", indicating strong learning about the workings of the system caused by reflection on this observed pattern.

Finally, the reveal of the prompt enhanced this process of finding patterns between input and outputs. The information provided by the prompt would not only confirm already found patterns but also equip the participants with new guidance to explore further. In most scenarios, except for participant three, the prompt included one or more names of artists or famous persons. In most cases, these names were unknown to participants, resulting in Google searches and comparing of the appearance of the mentioned persons or their artistic styles. All participants explicitly mentioned finding the revelation of the prompt after looking at the generated images a vital addition to the experience. They described the prompt as playing an essential part in their overall understanding of the system, as this offered them a tangible link between the camera input that was provided of themselves, and the five displayed images on the television screens. This illustrates the importance of the framework's Explanation phase when it comes to applying IIA to the field of XAI (Figure 7).

Additionally, a need for collaborative interaction was discovered. Whenever multiple participants were present during a user test, participants initially completed the interaction phases on their own, as was described in the protocol. However, after filling in the final part of the questionnaire, these groups of participants expressed the desire to re-enter the installation to experiment with different group configurations and input positions. This happened specifically in the first group, containing participants one, two, and three (Figure 19), and in the second group, where participants four and seven collaborated. It was observed that this shared experience enhanced both the Explanation and Reflection phase, as participants started to dissect and investigate the properties and links of the installation together, in an open and conversational manner.

**Fig. 19.** Participants forming a group during the user tests

## 6 Discussion

Drawing insights from XAI, XCC, and IIA, this study aims to bridge the gap between XAI and IIA. By leveraging the human-focused principles of XCC and the engaging nature of interactive art installations, new approaches, and opportunities can be envisioned for XAI that make complex AI models understandable and accessible. The IIA4XAI framework introduced in this study serves as a foundation for utilizing IIA as a design technique in the development of XAI systems and offers valuable guidance to designers and engineers who aim to involve non-expert users as part of their target audience. The framework bridges the gap between existing XAI systems that are primarily targeted at technical users with pre-existing knowledge of the field, and non-expert users that are currently left with no tools to help them understand complex AI technologies. Especially as this group of users is growing since more and more people are adopting new AI tools in their daily tasks. By highlighting the lack of understanding and engagement of this group of non-expert users with AI systems, this research addresses the limitations of current XAI approaches that primarily cater to technical users and advocates for a more inclusive and user-centered design approach. In addition, this study encourages the expansion of the goals and priorities within the field of XAI and stimulates broader discussions on the role of user experience in shaping the development of XAI systems.

The IIA4XAI framework was evaluated by means of a case study where the framework was applied for the development of an interactive installation based on AI image generation technology. Subsequently, the installation was tested with nine participants in the form of qualitative user tests. The participants of the user tests had a moderate level of tech-savviness and minimal prior experience with AI image generation systems. However, they exhibited a high level of openness towards adopting new technologies. After interacting with the installation, participants demonstrated an increase in general understanding, interest in learning about the technology, and engagement with it. On average, participants' general understanding increased by 108%, while interest and engagement increased by 31% and 79%, respectively. The average increase in interest of 31% was not as high as the increase in understanding and engagement. This could, however, be caused by the fact that the interest scores were already quite high, with an initial average of 6.4.

These findings suggest that the interactive installation effectively influenced this group of participants, improving their understanding of the artwork, and general interest in it, and enhancing their level of engagement. It could be possible that the choice of medium, in this case, the old CRT television sets, had an influence on the overall increase in engagement. This medium was deliberately chosen to invoke feelings of nostalgia, specifically for the group of participants above the age of 40, which is a proven tactic for enhancing engagement. To further investigate the influence of the chosen medium on the increase in engagement, the user tests should be repeated with the resulting images being displayed on a different medium, e.g. modern LCD monitors.

The results of the User Experience Questionnaire (UEQ+) showed that participants generally had positive experiences with the installation. The ratings for the scales Attractiveness, Stimulation, Novelty, and Quality of Content were all well above average. Some items falling into the scale of Intuitive Use, like the ease of use and logicality, were rated a bit lower. This could be caused by some complexities that had to do with the system still being in a prototype stage. However, since the current study only allowed for testing with nine participants, significant conclusions cannot be drawn. Therefore, it is recommended that for further research more extensive user tests are conducted with a bigger group of participants.

During the interaction, participants initially analyzed each image individually, looking for elements that reminded them of themselves. They recognized themselves based on various characteristics, such as appearance, posture, or mannerisms. Participants also started to identify patterns across the generated images, focusing on colors, compositions, and activities that appeared in multiple images. The reveal of the prompt, which included names of artists or famous persons, played a crucial role in participants' understanding of the system. It helped establish a tangible link between the input photograph and the generated images. Participants found the prompt to be an essential addition to the experience, leading to a deeper understanding of the system and motivating further exploration. This illustrates the difference between the phases of the framework applied to IIA (Figure 6) and the phases applied to IIA for XAI (Figure 7), and the importance of the Explanation phase.

Furthermore, collaborative interaction among participants was observed to enhance the experience. Participants expressed a desire to experiment with different group configurations and input positions, which facilitated open and conversational discussions during the explanation and reflection phases. Collaborative interaction is currently not part of the IIA4XAI framework, however, it could potentially be included in the framework as an additional component. In addition, the results of the user tests could have been influenced by the overall experimental setup of the user tests. The user tests took place in a lab inside the University, with the researcher being present in the space. The participants had to follow a protocol that involved instructions for the interaction, and reading and signing of multiple forms. All of these factors could have contributed to (unintended) biases or adjusted behavior of the participants.

For further research, the influence of the chosen medium on the increase in engagement should be investigated by repeating the user tests with the resulting images being displayed on a different medium, e.g. modern LCD monitors. Additionally, the prototype should be further developed and evaluated in a more natural setting, e.g. in a public environment or exhibition space, with complete freedom of interaction for the participants to minimize bias and adapted behaviour. Furthermore, to draw significant conclusions, the user tests should be conducted with a bigger group of (at least 30) participants. Finally, it is suggested that the introduction of the concept of collaborative interaction as a component in the framework is further explored.

## 7 Conclusion

In conclusion, this study addresses the existing gap between Explainable Artificial Intelligence (XAI) and non-expert users by introducing the IIA4XAI framework. The framework, informed by insights from XAI, XCC, and IIA, and the preliminary study, offers valuable guidance to designers and engineers who seek to involve non-expert users in the development of XAI systems. By incorporating human-centered principles and leveraging the engaging nature of interactive art installations, the framework enhances the understandability and accessibility of AI models. The evaluation of the IIA4XAI framework through a case study demonstrates its effectiveness in increasing the engagement of non-expert users with AI systems. The participants, who had minimal prior experience with the AI technology at issue: AI image generation systems, exhibited improved understanding, interest, and engagement after interacting with the installation. The displaying of prompts as a form of explanation and the affordance of collaborative interaction played crucial roles in encouraging discussion and reflection on the workings of the system. This research contributes to the field of XAI by advocating for a more inclusive and user-centered design approach and highlighting the importance of user experience in shaping the development of accessible AI systems for non-expert users. Ultimately, the IIA4XAI framework serves as a first step to equipping non-expert users with XAI tools to help them understand, gain interest in, and increase engagement with complex AI systems.

## 8 Acknowledgements

I would like to thank my advisors Dr. Rob Saunders and Dr. Peter van der Putten for guiding me during this research. Additionally, I would like to express gratitude to Fay Beening and Floor Stolk as members of my Committee of Critics. Thanks should also go to Mireille Stout for providing me with valuable feedback on my draft thesis. Lastly, I would like to thank all respondents in the preliminary study and participants in the user tests, this research would not have been possible without them.

## References

1. M. Van Lent, W. Fisher, and M. Mancuso, "An explainable artificial intelligence system for small-unit tactical behavior," in *Proceedings of the national conference on artificial intelligence*, pp. 900–907, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2004.
2. M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," *Communications of the ACM*, vol. 63, no. 1, pp. 68–77, 2019.
3. O. Loyola-Gonzalez, "Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view," *IEEE access*, vol. 7, pp. 154096–154113, 2019.
4. P. Hall and N. Gill, *An Introduction to Machine Learning Interpretability*. O'Reilly Media, Incorporated, second ed., 2018.
5. A. Rai, "Explainable ai: From black box to glass box," *Journal of the Academy of Marketing Science*, vol. 48, no. 1, pp. 137–141, 2020.
6. A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (xai)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.
7. M. U. Scherer, "Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies," *Harv. JL & Tech.*, vol. 29, p. 353, 2015.
8. G. Vilone and L. Longo, "Notions of explainability and evaluation approaches for explainable artificial intelligence," *Information Fusion*, vol. 76, pp. 89–106, 2021.
9. J. Bryson and A. Winfield, "Standardizing ethical design for artificial intelligence and autonomous systems," *Computer*, vol. 50, no. 5, pp. 116–119, 2017.
10. M. Turek, "Explainable ai, program information."
11. J. Gerlings, A. Shollo, and I. Constantiou, "Reviewing the need for explainable artificial intelligence (xai)," 01 2021.

12. M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should I trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pp. 1135–1144, 2016.
13. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 4765–4774, Curran Associates, Inc., 2017.
14. G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K.-R. Müller, "Layer-wise relevance propagation: an overview," *Explainable AI: interpreting, explaining and visualizing deep learning*, pp. 193–209, 2019.
15. S. Colton, G. A. Wiggins, *et al.*, "Computational creativity: The final frontier?," in *Ecai*, vol. 12, pp. 21–26, Montpelier, 2012.
16. N. M. Davis, "Human-computer co-creativity: Blending human and computational creativity," in *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2013.
17. M. T. Llano, M. d'Inverno, M. Yee-King, J. McCormack, A. Ilsar, A. Pease, and S. Colton, "Explainable computational creativity," *arXiv preprint arXiv:2205.05682*, 2022.
18. D. Rokeby, "The giver of names at the art gallery of windsor (2008)," *David Rokeby: The Giver of Names.*
19. R. Lozano-Hemmer and D. González, "Rafael lozano-hemmer: Pseudomatismos, muac museum, mexico city, mexico, 2016.,"
20. C. Bishop, *Installation art: A critical history.* London: Tate Publishing, 2005.
21. J. H. Reiss, *From margin to center: The spaces of installation art.* mit Press, 2001.
22. D. Mast, S. I. de Vries, J. Broekens, and F. J. Verbeek, "The participant journey map: Understanding the design of interactive augmented play spaces," *Frontiers in computer science*, p. 45, 2021.
23. E. Huhtamo, "Trouble at the interface, or the identity crisis of interactive art," 2004.
24. R. Kluszczynski, "Strategies of interactive art," *Journal of Aesthetics & Culture*, vol. 2, no. 1, p. 5525, 2010.
25. K. Kwastek, *Aesthetics of interaction in digital art.* Mit Press, 2013.
26. K. Feingold, "Ou: Interactivity as divination as vending machine," *Leonardo*, vol. 28, no. 5, pp. 399–402, 1995.
27. T. Wildschut, C. Sedikides, J. Arndt, and C. Routledge, "Nostalgia: content, triggers, functions.," *Journal of personality and social psychology*, vol. 91, no. 5, p. 975, 2006.
28. M. Schrepp and J. Thomaschewski, "A modular extension of the user experience questionnaire," *UEQ+*, 2019.

# A Supplementary Figures

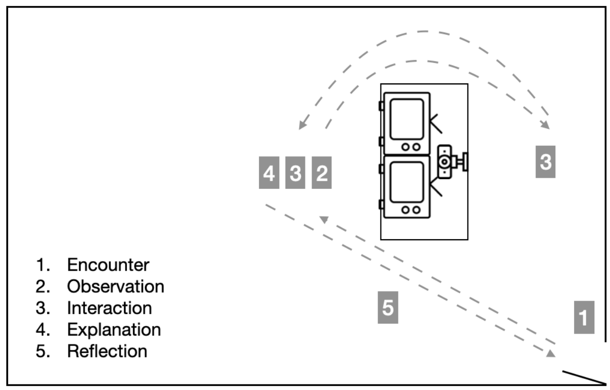


**Fig. 1.** Schematic overview of the phases in the installation space

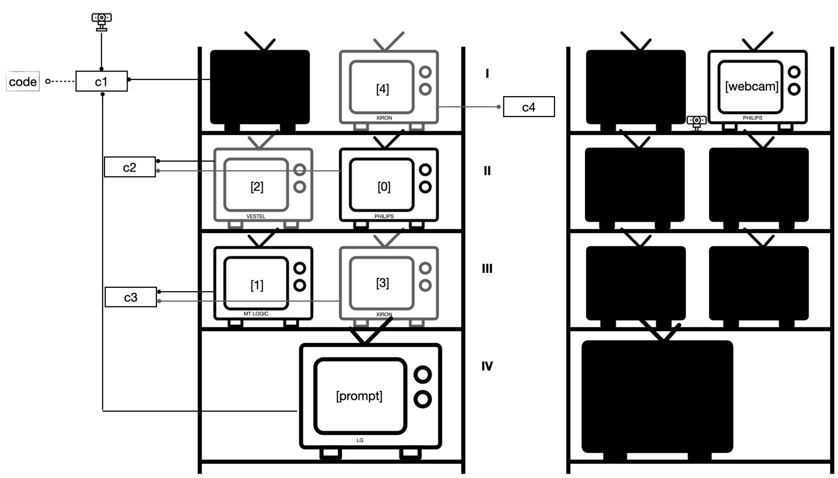


**Fig. 2.** Schematic overview of installation setup

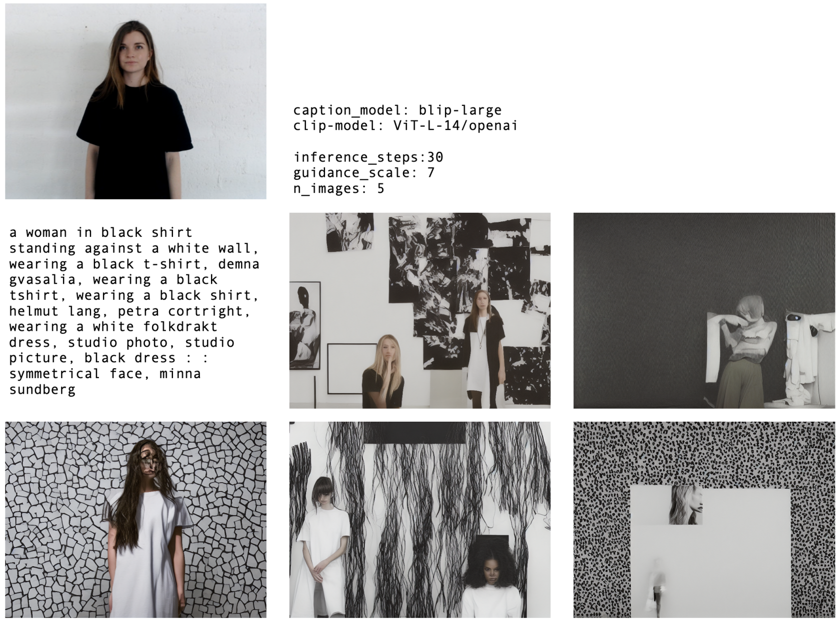


**Fig. 3.** Example of the output of the generation process, from top left to bottom right: input image, model parameters, generated prompt, x5 generated images

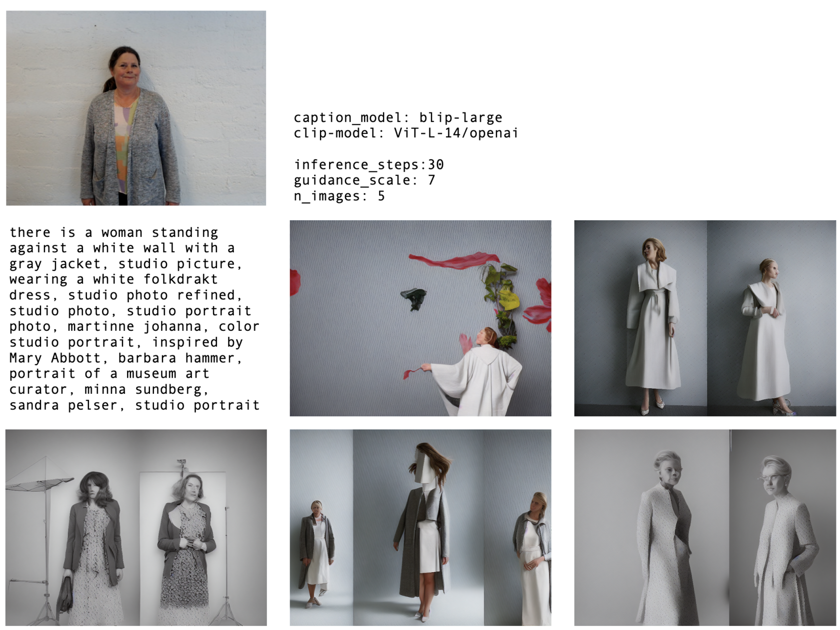


**Fig. 4.** Example of the output of the generation process, from top left to bottom right: input image, model parameters, generated prompt, x5 generated images