



Investigating AI-Ethics through Necromancy

An Interdisciplinary Exploration of Uncertainties in
Performative Technologies as Curated AI Personas

Master's thesis in Interaction Design and Software Engineering

MAX HAGMAN & SALLY KALM

MASTER'S THESIS 2026

Investigating AI-Ethics through Necromancy

An Interdisciplinary Exploration of Uncertainties in
Performative Technologies as Curated AI Personas

MAX HAGMAN, SALLY KALM



UNIVERSITY OF
GOTHENBURG



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2026

Investigating AI-Ethics through Necromancy
An Interdisciplinary Exploration of Uncertainties in
Performative Technologies as Curated AI Personas
MAX HAGMAN & SALLY KALM

© MAX HAGMAN & SALLY KALM, 2026.

Supervisor: Jasmina Maric, Interaction Design and Software Engineering
Advisor: Monika Fleischmann and Wolfgang Strauss
Examiner: Michael Heron, Interaction Design and Software Engineering

Master's Thesis 2026
Department of Interaction Design and Software Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: User interface containing four AI-philosophers in an exhibition setting

Typeset in L^AT_EX
Gothenburg, Sweden 2026

Investigating AI-Ethics through Necromancy
An Interdisciplinary Exploration of Uncertainties in
Performative Technologies as Curated AI Personas
MAX HAGMAN & SALLY KALM
Department of Interaction Design and Software Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

This thesis was driven by its uncertain nature, where emerging AI technologies was investigated through Research through Design, Speculative Design and an artistic process. This perspective gave way to exploring solutions that would otherwise have been overlooked as unconventional. The project resulted in an interactive and performative art show, where a debate between four deceased media-philosophers was simulated through the use of Large Language Models. The show's aesthetics, atmosphere and content all contributed to the User Experience. It was developed iteratively with formative and summative evaluations in form of interviews and surveys, with continuous on-site testing. The critical reflections on AI use-cases resulted in several ethical insights: there is a need to design for responsibility and accountability of generated outcome; there is power in orchestration, as the persons imitated with AI gain importance; there is a need for transparency in training data and agenda; and finally, anthropomorphism contributes to building relationships between humans and machines. Moreover, this thesis also reflects on the possibility of generating new knowledge, and concludes that personified AIs might be a tool to gain lost perspectives on current world states, but will not necessarily generate new knowledge with the current technological limitation. Lastly, This project calls for further research into charismatic AIs and guidelines for social chat-bots, as well as further research into integrity and consent when it comes to simulating people through AI technology.

Keywords: AI, ethics, interaction design, computer science, engineering, art.

Acknowledgements

"This world is changing... We decide into what."

– Kai'Sa, *League of Legends*

"Those who do not know the danger of wielding power will, before long, be ruled by it."

– Lanayru, *The Legend of Zelda: Twilight Princess*

We would like to express our sincere gratitude to everyone who contributed to this project. Whether you were involved in development, participated in our user testing, or joined us for the final show, thank you for your invaluable time and effort.

We would like to thank our supervisor and project lead, Jasmina Maric, for her feedback, positivity and trust throughout this uncertain project. This work would not have been possible without her expertise and direction. In addition, we would also like to thank our Examiner Michael Heron.

We would also like to extend our thanks to Monika Fleischmann and Wolfgang Strauss, for their creative vision, motivation and willingness to fly out two times to Gothenburg to make this project possible.

In addition, we would like to express our gratitude to Rasti Mhealden Tengman for his endless dedication and late-night hours in an effort to make the system work. Without him, this show simply would not have been able to happen. Moreover, we are grateful to Scionova AB for hosting us and providing guidance.

Finally, we would like to thank our dear friends who worked alongside us during this entire project. Thank you for the coffee-breaks, the thrilling rounds of *Ordel*, and the overall supportive and motivating environment.

Max Hagman & Sally Kalm, Gothenburg, 2026-06-12

Contents

List of Figures	xiii
List of Tables	xv
1 Introduction	1
1.1 Research Question	2
1.2 Aim & Goal	2
2 Background	3
2.1 Related work	3
2.1.1 Home of the Brain	3
2.1.2 Thinking Space	4
2.1.3 Previously curated AI-Philosophers	5
2.2 Project Brief	5
2.2.1 Concept Description	5
2.2.2 Stakeholders & Roles	7
3 Theory	9
3.1 Design Theories	9
3.1.1 Interaction Design & Human Computer Interaction	9
3.1.2 Research through Design & Designing for Value	10
3.1.3 User Interface Design	12
3.2 Designing an Interactive Exhibition	13
3.3 Designing for Large Language Models	13
3.4 Ethical and Social Considerations	14
3.5 Designing for Uncertainty	15
4 Methodology	17
4.1 Converging Views	19
4.2 Ideation	21
4.3 Early Formative User Tests	22
4.4 Concept Development	23
4.5 Audience Tests and Iterative Refinement	24
4.6 Presentation of finished Design	25
5 Results	27

5.1	Defining the Project Scope	27
5.1.1	Magic, Wow and Effect	28
5.1.2	Story Mapping and System Description	28
5.1.3	User Journeys	30
5.2	Ideation & Initial Concepts	32
5.2.1	Design Decisions	32
5.2.2	Concept Iterations	33
5.2.3	Group Interviews	34
5.3	Prototype Implementation & Development	37
5.3.1	Art Assets Development	37
5.3.2	Front-end Development	39
5.3.3	Supporting Back-end Development	40
5.4	Audience Evaluations	41
5.4.1	Wisdome User Test 1	41
5.4.2	Survey Results from Wisdome User Test 1	43
5.4.3	Thematic Analysis of Wisdome User Test 1	44
5.4.4	System Design Iterations	47
5.4.5	Wisdome User Test 2	49
5.4.6	Survey Results from Wisdome User Test 2	51
5.4.7	Final Design Iterations	52
5.5	Final System Design & Functions	53
5.5.1	Front-end Interfaces	54
5.5.2	Environment Controls	57
5.5.3	Human-to-AI Interaction	57
5.6	Final Showcase	58
5.6.1	Survey Results from the Final Showcase	60
6	Discussion	67
6.1	<i>Magic</i> , the gathering of Value	67
6.2	Necromancy as an Interaction design practice	69
6.3	The Thesis is now a Kitchen Appliance	72
6.4	Stakeholder meetings, a Barbecue of Thoughts	72
6.5	Art over Accessibility	73
6.6	Answer to the Ultimate Question of Life, the Universe, and Everything	75
6.6.1	Responsibility of Generated Outcome	75
6.6.2	Orchestration and Transparency	75
6.6.3	Personification; Consent, Trust and Relationships	76
7	Conclusion	79
	Bibliography	81
A	Appendix	I
A.1	Design Criteria	I
A.2	Functions and corresponding Solutions to be Combined	III
A.3	Initial Concepts for Scenario 2: Wisdome	V
A.4	Initial Concepts for Scenario 1: Installation	VII

A.5	Semi-structured interview guide	VIII
A.6	Survey Questionnaire Template 1	X
A.7	Survey Questionnaire Template 2	XV
A.8	Survey Questionnaire Template 3	XVIII
A.9	State-changing Functions & updated Story Mapping	XXII

List of Figures

2.1	Images from Monika Fleischmann and Wolfgang Strauss' artwork <i>Home of the Brain</i> (1990-1992) [9], presented in the FLUX interview [10]	4
3.1	Dealing with non-routine design problems, by Daalhuizen et al. [47] .	16
3.2	Uncertainties as Generative Resources in RtD, by Epp et. al [48] . . .	16
4.1	Illustrating how exploring uncertainties can reveal new ones, creating an unpredictable path and process	17
4.2	Overview of the iterative approach-loop	18
4.3	Visual process describing all design iterations	19
4.4	Mindmapping value-oriented goals and ideation process	20
4.5	First Wisdome visit	21
4.6	Stakeholder meetings and consultation	23
5.1	Story mapping session with post-its	28
5.2	Initial human-machine-system	29
5.3	User Journey of <i>Scenario 1: Installation</i>	30
5.4	User Journey of <i>Scenario 2: Wisdome</i>	31
5.5	Visual concept 1: text-based UI, and background visuals in dome-setting	33
5.6	Visual concept 2: text-based UI, and background visuals in a dome-setting	33
5.7	Concept 1: Portraits in a row with decorations, dither style portraits, non-static user question, minimal text output, bottom-to-top flowing text	34
5.8	Concept 2: single AI-philosopher portrait, ASCII style portraits, static user question, chunks of text, top-to-bottom flowing text	34
5.9	Concept 3: Single AI-philosopher portrait with decorations, ASCII style portraits, non-static user question, chunks of text, top-to-bottom flowing text	34
5.10	Thematic analysis of group interviews	35
5.11	The four style alternatives of AI-Flusser's portrait	38
5.12	360 fulldome format background visual and mock-implementation . .	38
5.13	The UI projected in Wisdome with overlapping background visuals .	39

5.14	Simplified human-machine system, where pink describes the audience interaction, orange the back-end system process and the UI to the top right	41
5.15	Images from the first user test in Wisdome	42
5.16	General position toward generative AI tools	43
5.17	Results of rating the experience based on statements	44
5.18	Survey answers regarding accountability	44
5.19	Thematic Analysis of back-end themes	44
5.20	Thematic Analysis of front-end themes	46
5.21	Images from the second user test in Wisdome	50
5.22	The final visual description of the human-machine system, the pink color represents controls not visible to the audience, the green color represents controls visible to the audience, and the blue color represent the back-end.	53
5.23	The final interface projected in Wisdome while AI-Virilio is speaking	54
5.24	Portraits and user question while AI-Weibel is speaking, with low distance	55
5.25	Key frames of AI-Flussers normal, "calm" talking animation	55
5.26	Key frames from AI-Flussers low distance "extreme" talking animation	55
5.27	All frames in AI-Flussers thinking animation	56
5.28	The interface of the control panel	56
5.29	Boot-up and credit screens	57
5.30	Background visuals together with the credit scene in Wisdome	58
5.31	The AI-philosophers in Wisdome discussing the final question, AI-Virilio is currently speaking.	59
5.32	General position toward generative AI tools	61
5.33	Rating experience of the AI-philosophers based on these statements .	62
5.34	Rating experience of the interaction based on these statements	63
5.35	Survey answers regarding accountability	65
A.3.1	Visual concept 1: text-based UI, and background visuals in dome-setting	V
A.3.2	Visual concept 2: text-based UI, and background visuals in a dome-setting	V
A.3.3	Figure (a), (b) describes interaction concept 1:, record and release mobile UI screens. Figure (c), (d) describes interaction concept 2: voting and asking questions mobile UI screens	VI
A.4.1	Installation concept 1: Input screen, and exhibition with output screen	VII
A.4.2	Installation concept 2: Input screen, and exhibition with output screen	VII

List of Tables

5.1	Group Interview participants	35
5.2	Wisdom User Test 2 participants	49
A.1	New state changing functions & their relevance in different use-states	XXII

1

Introduction

Life is full of uncertainties, be it global pandemics, relationships, or even what to eat for dinner. Dealing with uncertainties, personal or professional, can be a stressful and daunting task, forcing you to make decisions based on parameters that are out of your control. So, what can we be sure off? By confidently stepping head-first into the unknown, we do not know where we will end up. While this initially may seem like an undesirable tradeoff, the alternative, a path already laid out, leaves little room for curiosity, exploration or discovery. Embracing the uncertain, will leave you with experiences that could never have been planned for.

The same goes for academic research. Investigating highly uncertain fields and embracing openness, learning and failing, will ultimately generate unexpected results. Frameworks that adopt this mentality are Research through design and Speculative design [1], [2], described as pathways to look into unforeseen futures. Through exaggeration and satire, their purpose is to challenge the perception of the current state of the art, and generate knowledge through the act of making.

This thesis takes on a speculative and exploratory approach to an uncertain concept. With the rise and effectiveness of *Artificial Intelligence* (AI) within multiple sectors and in daily life, there is an implication of high trust in this new technology [3]. At the same time, there are many factors in the design of *Generative AI* and *Large Language Models* (LLMs) that negatively affect the *User Experience* (UX). For example, they are prone to being biased which leads to discriminatory outputs [4], [5] and show concerning issues in being highly sycophantic [6].

Indicating that due to rapid advancement, the framework for guiding its morality is still catching up. To guide development, the European Commission's ALTAI framework establishes seven core guidelines for trustworthy AI: human agency, technical safety, privacy, transparency, fairness, societal well-being, and accountability [7], [8].

Thus, researching this field in a provocative and exploratory way might give insights into how these guidelines function and what is lacking. Rather than evaluating AI ethics in a controlled lab setting through a rigid framework, this thesis utilizes artistic artifacts as deceased media philosophers are simulated through AI technologies to test human-ai interaction in a entirely new context. Combining artistic tools with scientific rigorous research, this study investigates the socio-technical issues arising from the rapid rise of AI and envisions future use cases in order to ultimately challenge how AI is perceived.

In this study, Generative AI refers to systems designed to create content, such as text, images, or audio, by learning from existing data patterns. The term AI refers broadly to the public perception and discourse surrounding artificial intelligence.

This project used generative AI during development for the following purposes: generating code for the front-end interface; generating code to program background visuals; to understand code structure and back-end functionalities; and grammar aid.

1.1 Research Question

Due to the fast advancement in AI technology and its implications on socio-technical challenges, we are phased with issues regarding the current state of Generative AI models, especially its emotional and affirmative characteristics, and its ability to make users trust its output as original. This could be the start of outsourcing human critical thought. To further understand this emerging technology and its ethical implications on society, this study presents the following research question:

What insights to AI-ethics emerge in the interaction with LLMs simulating conversation with deceased media-philosophers, within an artistic process?

1.2 Aim & Goal

The aim was to take part in shaping emerging AI technology, by using art and design frameworks as a critical tool. By engaging users, the aim was to spark reflection about AI-technologies, and to let users be apart of the discourse of the future of AI.

The goal was to, through an exploratory and speculative lens, contribute to existing research on AI-ethics and design an artistic performance showcasing an interactive conversation between LLM philosopher personas. This encompasses the design and implementation of an interface and visual elements supporting the interaction between the users and the LLMs.

To answer the research question and achieve the aim and goal, this Master's Thesis was born. To unravel the possibilities held within the uncertainties, the study begins with presenting the background context for the thesis work. This establishes earlier artistic interventions made in the intersection of art, AI and ethics. This is followed by the introduction of the stakeholders that took part in the project, as well as the original artwork *Home of the Brain* (1990-1992) [9], from which the interactive artwork *Mind Circuits*, developed in this project, conceptually emerged. The thesis then presents the theoretical and methodological frameworks that laid the foundation of the work, as well as how these frameworks were applied in practice. Lastly, the results are presented, following a discussion based on the ethical implications of AI-human interaction through the lens of an artistic context, as well as its connections to exploratory design practices.

2

Background

This chapter presents the background context for the thesis work. The following sections are *2.1 Related work*, which describes predecessors and inspirations, and *2.2 Project brief*, which illustrates the initial idea and the associated stakeholders.

2.1 Related work

This section covers the original artwork *Home of the Brain* and its connection to the concept of *Thinking Spaces*. Furthermore, it introduces previously curated AI representations of prominent philosophers.

2.1.1 Home of the Brain

The interactive artwork *Mind Circuits*, developed as part of this thesis study, was based on an idea of Monika Fleischmann and Wolfgang Strauss (MF & WS) to recreate their artwork *Home of the Brain* using modern day AI technologies. Home of the Brain (1990-1992) was a *Virtual Reality* (VR) installation by MF & WS [9]. The visitor would navigate the VR space using a control called a "data glove", and other visitors could experience this on a large projector screen. The work was both a showcase of what was at the time a new medium and an invitation into the minds of four prominent philosophers. Each philosopher was represented through "houses", with bold colors and abstract shapes, as well as a more realistic representation, with real images and voices. The houses showcased an audiovisual experience, movement became a way of thinking through the data-glove, see figure 2.1. Altogether, Home of the Brain carries the discussion of media into the media itself.

The original project was based on the ideas of Joseph Weizenbaum, Marvin Minsky, Paul Virilio and Vilém Flusser. [9]. The philosophers conflicting ideas about modern technology and digitization could through this work be experienced in the virtual world, in a way diving into their minds. The houses they represented were: Minsky - Utopia, having a positive and optimistic view on future AI technology; Weizenbaum - Hope, being a mindful yet hopeful critic of AI from an ethical standpoint, after developing the first ever chat-bot; Flusser - Adventure, who presented new ideas and viewpoints about the human-technology systems through the lens of the apparatus; Virilio - Catastrophe, being a voice of concern that the speed and acceleration of technological development will leave humans behind.

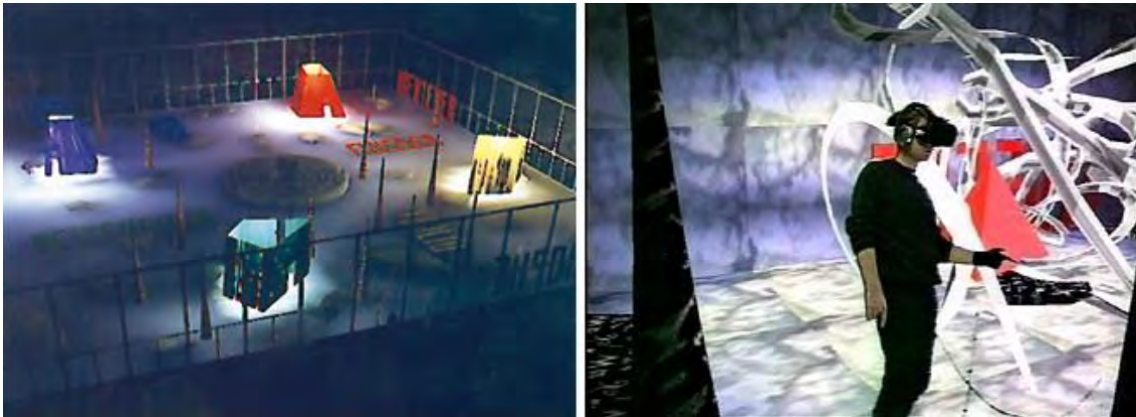


Figure 2.1: Images from Monika Fleischmann and Wolfgang Strauss' artwork *Home of the Brain* (1990-1992) [9], presented in the FLUX interview [10]

From an interview with the artists [10], the interviewer mentions that *Home of the Brain* is one of the artists' most intriguing and talked about works. He continues to describe it as being frequently discussed in books covering the history of interactive art and is generally considered one of the most important works of its decade. He highlights the interaction as being the core part of the concept, and through the interaction, it facilitates deeper thought within the subject the work touches, such as digital futures and virtual spaces.

2.1.2 Thinking Space

The concept of the Thinking Space, or *Denkraum* originates from cultural scientist Aby Warburg's work from the 1920s, which describes a space where new knowledge is extracted by arranging and rearranging visual and textual information [11]. In the context of interactive media art, MF & WS redefine the thinking space as a "space filled with data". The authors also mention that Thinking Spaces functions as a medium where reflection is facilitated through active user participation.

As Fleischmann and Strauss discuss, works in Thinking Spaces are created to make people think critically about the intersection of society and technology. By placing participants in immersive, participatory, and mixed-reality environments, it encourages participants to think critically about what they are experiencing by interacting with the material [11]. The authors describes that the works utilize performative interfaces that shift the traditional user experience from "What you see is what you get" to "What you get is what you have not seen before". Ultimately, works in Thinking spaces raise an important question: as automation improves and takes over, will humans continue to think and act responsibly, or will machines take over? [11]

Thus, *Home of the Brain* is a work that fits the concept of a Thinking Space. The environment in which the user is transported is far beyond familiar. Traversing this digital landscape and absorbing its data, it evokes curiosity and discovery, which invites philosophical reflection on the very digital landscape being traversed.

2.1.3 Previously curated AI-Philosophers

Simulation personas through artificial intelligence has come a long way since the early eras of the internet. In the late 90's, students Corey Kosak and Darrell Kindred at Carnegie Mellon University released their website called *Forum2000* where Quadratic Search Algorithms powered an early AI-chatbot [12]. This chat-bot had different personas who would answer based on their personality filters called SOMAD. The SOMADs were not necessarily philosophers, there was a neutral personality called "the cube", a satirical representation of the "thoughts of god" called "the Oracle", and many personas were based on real life people like Bill Gates. While the conversations generated were short and sometimes incoherent, this project was a successful early experiment in curated AI personas.

Since then, a lot has happened. In 2020, the interactive VR installation *CHOM5KY vs CHOMSKY* [13], that allows you to communicate with an AI model of philosopher and linguist Noam Chomsky, premiered at the Sundance Film Festival. Rodriguez, project lead, stresses the importance of using artistic expression to demystify AI and encourage public participation in the discourse [13]. Using humor, collaboration, and exploration, this project aimed to challenge the mainstream view of AI by clarifying general concepts in a collaborative and interactive way.

Schwitzgebel et al. [14] created a LLM of philosopher and cognitive scientist Daniel Denett in 2024, to test if philosophy, like other writing oriented fields of work, are endangered by AI. They found that you could not reliably tell apart the output the AI model produced from the real philosophical statements that Denett provided. However, the model was not always accurate when imitating Denett. The authors stress the risk of over-reliance on such models, since they are prone to inaccuracies, which some inexperienced users missed out on [14]. Both Rodriguez [13], and Schwitzgebel et al. [14] stress important topics such as transparency and trust in AI models. They also speculate on a future scenario that encapsulates real people into digital avatars, capable of "thinking" and "existing" on its own.

Kuenen [15] highlights the importance of the aesthetic experience in interaction design, and notes the close connection between emotional aspects of interaction and cognitive science [15]. In the swift emergence of generative AI, future use-cases are uncertain and speculative design through the artistic medium might bring further insight into user interaction and perception of AI.

2.2 Project Brief

This section covers the project brief. This includes the initial *Concept Description*, containing directions for the project start as well as descriptions of the philosophers, and the relevant *Stakeholders* associated with the project and their roles.

2.2.1 Concept Description

The new artwork is called *Mind Circuits* and is an interactive live museum experience, connecting historical ideas with contemporary AI systems. According to MF & WS

it can be described as a performative Thinking Space: a staged space to think about AI, ethics, media and society. The Concept Description included directions that the audiences should be able to interact with four LLMs, as AI-philosopher personas, actively shaping the discussion through questions and provocations. The concept originated as a remake of Home of the Brain using modern day AI-technologies. Therefore, aesthetic elements such as the color coding were important as a homage to the Home of the Brain. *Mind Circuits* aims to encourage reflection on technology, ethics, and society, and visitors may leave feeling intrigued, unsettled, amused, or challenged.

The target audience is broad, focusing on people interested in interactive experiences and AI/emerging technologies, with proficiency in English. However, participants will be invited regardless of interest in interactive AI experiences and English proficiency. These are some concepts from the brief that were relevant in development and ideation:

- The LLMs should be aware of being artificial intelligence
- They should not claim authenticity or identity with the historical individuals
- The LLMs should be informed of the modern world
- There should be both AI-to-AI and Human-to-AI dialogue
- The audience can ask any question, and should be able to interrupt any time
- The system does not aim to provide answers

The brief included directions on the LLMs training data, being published texts, lectures, interviews and public statements. To capture the knowledge and rhetorical styles of Joseph Weizenbaum, Vilém Flusser, Paul Virilio, and Peter Weibel. The first three mentioned were chosen based on the artwork Home of the Brain, however, due to ethical considerations, Marvin Minsky was replaced by Peter Weibel.

Having a personal connection with each of the philosophers, MF & WS, shared a written statement of each persons history and personality, to support the LLM's training data. A summary of their description of the philosophers follows:

Joseph Weizenbaum (1923-2008)

Weizenbaum was a German-American computer scientist and one of the first critics of artificial intelligence, following the empathic reactions to his invention of the conversational program *ELIZA*. Fleeing Nazi Germany in 1936, he later became Professor of Computer Science at MIT. Core beliefs: Computers calculate, humans judge; Moral responsibility cannot be delegated; Capability does not imply permission; Some tasks must remain human.

Vilém Flusser (1920-1991)

Flusser was a Czech-born philosopher who also fled the Nazi occupation, losing most of his family in the Holocaust. After emigrating to Brazil, he developed a philosophy of communication largely outside formal academic institutions. Core beliefs: Media

shape consciousness more than content; Apparatuses program behavior; Freedom exists only through play against systems; Technology is ambivalent, not moral.

Paul Virilio (1932-2018)

Virilio was a French theorist whose childhood under the Nazi occupation and later work in architecture shaped his analysis of speed, war, and technological disaster. Core beliefs: Speed is the dominant political force; Every technology invents its accident; Automation erases human presence; Progress produces blindness.

Peter Weibel (1944-2023)

Weibel was an Austrian media theorist, artist, and curator, best known for his leadership at the Center for Art and Media Karlsruhe. He treated institutions as apparatuses and moderation as intervention. Core beliefs: Institutions distribute power; Neutrality conceals domination; Conflict produces knowledge.

2.2.2 Stakeholders & Roles

Monika Fleischmann and Wolfgang Strauss are digital media artists working with interactive and immersive virtual landscapes as Thinking Spaces [16]. MF & WS are the original creators of Home of the Brain, and were the principal authors of this artistic project. They were responsible for the artistic concept and general coherence.

Jasmina Maric is a Senior Lecturer at Interaction Design and Software Engineering at Chalmers University of Technology. She was the project lead, tasked with coordinating different stakeholders, steered the conceptual direction and supervised the interaction design development.

Rasti Mhealden Tengman is a master student from Chalmers University of Technology. He was responsible for the back-end engineering and development of the LLMs.

Scionova AB is an IT consultancy company focused on AI, Internet of Things and integrated systems. Scionova AB provided technical expertise, supervision and validation for the back-end development of the project.

Our roles focused on designing, developing and testing. This means crafting digital spaces, creating human-AI interaction concepts, as well as interface design and implementation. Additionally, we were responsible for planning and structuring the final event, as well as collect and analyze data, and perform user tests throughout the development process.

2. Background

3

Theory

The following sections describe the theoretical and conceptual frameworks which encompasses the intended systems, concepts and theories that the project is based on. The sections are: *3.1 Design Theories*, *3.2 Designing an Interactive Exhibition*, *3.3 Designing for Large Language Models*, *3.4 Ethical and Social Considerations* and *3.5 Designing for Uncertainty*.

3.1 Design Theories

This section will introduce concepts and methodologies related to design studies, both conceptual and theoretical. These design fields provided the basic framework and approaches in development.

3.1.1 Interaction Design & Human Computer Interaction

One of the main focuses in this project is audience interaction, and therefore *Interaction Design* (ID) and *Human Computer Interaction* (HCI) are two main fields of study which informed the design methodology and approach. ID, as described by Sharp et al. [17], is the design of the interaction between users and products, and by improving the experience, also improving the everyday life. The authors also describes the HCI field as one that largely overlaps with ID, differentiating them through describing HCI as having a greater focus on *Usability* within computing systems. ID & HCI offer multiple different frameworks and concepts for both practice and discussion. Applying these frameworks is essential in order to gain knowledge and explore topics of interest within these fields. These are both user-centered design fields. By designing the experience, you are designing for the user.

User Experience (UX) refers to the behavior of a product or system, when users from the real world are interacting with it [17]. The authors describe that ID and HCI share UX as a core pillar of design, since it is all about the people whom you are designing for. Thus, the User Experience is embedded in the product or system itself. Therefore, it is the real world users who gets to decide what good UX is. Moreover, Sharp et al. [17] mentions the broad scope of UX, which encompasses many design aspects such as *usability, functionality, aesthetics, content, look and feel* and *emotional appeal*.

Therefore, designers must be able to put themselves in the user's position to truly understand their needs and behaviors [17]. There are multiple methods for this specific UX related problem, which will be explored in chapter 4. Methodology, but more broadly, understanding the user requires the designer to practice empathy and engage on a deeper level with the user's real experiences. By using imagination as a tool to anticipate how a user will feel and react, the design process then becomes a journey of discovery for the designer [17].

A framework for facilitating this exploration is *Storytelling*, which can be helpful for many different purposes. There are approaches within Storytelling to aid in interdisciplinary communication, understanding the perspective of the user as well as a way to generate user engagement in the final design [18]. Furthermore, the author means that Storytelling can be used in prototyping and formative evaluations to help cover a wide range of the product experience.

Understanding users in HCI is a not straight forward process, and resolving issues in this field is often difficult as there is rarely a single, definitive solution. This aligns with the concept of *Wicked Problems* which refers to problems characterized by the absence of a single correct answer [19]. The author describes that the solutions to problems of such complexity cannot be correct or false, instead they can only be judged by being good or bad. Moreover, the subjectiveness of the good and bad has to be made explicit through a process of objectification so that the basis of one's judgment is communicated to others [19].

3.1.2 Research through Design & Designing for Value

Lockwood [20] suggests that value in design is often synonymous with economic gain, either through increasing user satisfaction, improving time to market processes or generating a solid return of investment. However, Lockwood stresses that designing for value can also emerge from the knowledge generated by the process itself. This, in turn, shifts the definition of value to centrally be about the rich insights and understanding gained through the act of making.

This project will include different approaches of conducting research. *Research through Design* (RtD) focuses on designing and evaluating artifacts as the primary method for gaining new data and knowledge about design related issues as described by Godin and Zahedi [1]. The authors claim that, in many ways, RtD can be seen as a more hands-on research approach, where the research question is as important as the process and outcome of the design itself. Godin and Zahedi [1] mean that there are ongoing challenges in justifying the validity of RtD, as critics argue it is not a traditional or proper research method. However, the authors stress that if the project works and the produced artifact is acceptable, the knowledge generated through its creation is in turn valid, as the ultimate goal of RtD is to produce knowledge and understanding. This knowledge results directly from the making process and, as stated earlier, is fully embodied in the created artifact.

Several design methodologies resonate with the core concepts of research through design, each shifting its focus from commercial gain to the production of knowledge. In these frameworks, the final outcome is unpredictable, serving primarily as a way to inquire new information about areas of research, or questioning the current state of said areas. Examples of this approach are: *concept-driven design*, *speculative design* and *critical design*.

Concept driven design can, according to Stolterman and Wiberg [21], manifest as the design process being conceptual and theoretical, tied to RtD as a conceptual exploration through development of artifacts. Furthermore, Stolterman and Wiberg [21] also highlight that the end result will be optimized for an idea, concept or theory rather than a problem or user. As a design principle where the core concept is fundamentally different from any in exciting products, the exploratory design methodology results in innovative solutions [22].

Adjacent to Concept driven design, is *Speculative design*, which also operates in the exploratory and unknown. Speculative design focuses on designing for speculative futures, and through that, is an active critique on current technological paradigms, as described by Malpass [2]. It questions the role of new technology, often told through a satirical and dystopian lens. Malpass compares speculative design with *critical design*, which both stems from designing in a way that questions the role of products and technology, both present and future. Furthermore, the author stresses that speculative and critical design uses satirical elements to challenge existing paradigms. However, critical design focuses on the present societal implications of existing design objects and technology.

Closely related to RtD lies *Artistic Research*. Artistic Research describes the bridge between science and art [23], [24]. Similarly to other scientifically based research practices, Artistic Research embodies a search for knowledge that aims to understand the world around us [23]. However, artistic Research has an emotional and aesthetic approach, with culture, perspectives and traditions in focus. Through art, because of its abstract nature, problems of high complexity can be investigated.[24]. Here, Klein argue that insights from Artistic Research is inquired through the creation of emotional and sensory experiences, resulting in embodied, "felt knowledge". In this way, RtD and Artistic research overlap in being hands-on research frameworks, where both has the act of making as a central role driving research forward [25].

The core of this thesis project was the development of an art piece that creates emotional and critical reflections within the audience, and in that sense, Artistic Research might be more representative for the theoretical framework that was followed. However, the conceptual idea for the artwork was already determined and therefore not a part of development. Instead, the development, design and evaluation of the interaction, system and aesthetics were the focus of this thesis. Due to the highly technical parameters of the artwork and the UX focus, the choice to view this project through the lens of Research through Design was made.

3.1.3 User Interface Design

A *User Interface* (UI) is a tool that allows a user to interact with a system [17]. Therefore, in UI design, it is important that the presentation, look, and feel of the system and its elements support the user in interacting with the system. However, Roth [26] describes UI design as more than the look and feel, as "it sets the tone for the entire user experience, from setting the mood and evoking an appropriate emotional response through improving usability and subjective satisfaction." He continues with describing the process of designing for user interfaces as being highly creative, and that the iterative process of creating valuable user interfaces is necessary in achieving coherency and actual usefulness in the design.

When creating visual interface designs there are several usability heuristics to consider as described by Nielsen [27]. While all are significant in achieving quality user experience, this study focuses on four principles from the authors collection that are most relevant to the application:

Visibility of System Status: The interface should inform users about what is currently happening, by providing feedback when a task is performed within a reasonable time-frame.

Aesthetic and Minimalist Design: The interface should only contain information that is contextual or relevant to the action that the user is taking. Extra elements risks over-complicating and thus reducing the value of all other important elements.

Consistency and Standards: The elements in an interface should be consistent with industry conventions and standards, so that the user understand the interaction based on previous experience. This ties in with the concept of *Mental Models*, which Norman [28] describes as personal cognitive images of how things in the real world are supposed to work.

Error Prevention: Error Prevention in user interfaces is about minimizing errors, making them invisible to the user, or ultimately, ensuring the user feels comfortable resolving them.

In the context of usability heuristics, the design and placement of interface elements directly influence how intuitively a user comprehends the system. *Visual hierarchy*, in regards to human-computer interaction, concerns the order in which users read information of a UI. Thus, how and where to place information to create a structure that highlights the most important elements [29]. The author mentions that using correct typography, color, and sizes for visual elements is important to create a logical flow of information which in turn improves the visual hierarchy.

Design and placement not only affects how information is perceived, but also informs the user which action should be taken. Too much information at a given time can affect the users *Cognitive load*, which refers to the amount of mental strain on the working memory at any given time [30]. Cognitive load, in regards to human computer interaction, is described by Hartson [31] as whether or not users are able to complete tasks based on the amount elements and thus, the amount of choices to be made when interacting with a user interface.

3.2 Designing an Interactive Exhibition

Many articles that discuss frameworks for designing interactive exhibitions focus on exhibitions for science museums, which are often curated by, or in close collaboration with the museum staff [32], [33], [34]. The User Experience is in focus, as the interaction is the driving force in the response of the visitors. Therefore, there exists a cause-effect relationship between the user response and the change in the exhibition [33].

There are similarities to designing interactive exhibitions for both art installations and science museums. In contemporary and new media art exhibitions, the audience almost plays the role of co-creator, as their presence reflects and influences the perception of the artwork [35], [36].

User tests are essential to good UX design, and in the context of museum exhibitions, Bodin and Duvanå [32] conclude that user testing is incredibly difficult. This is largely because there are rarely opportunities during the design process to perform user tests in the exhibition context [32]. In addition, their research points to the lack of formative testing. This means that when the exhibition is installed and the possibility of performing user tests emerges, the design is usually already finished; resulting in the user tests being summative rather than formative.

Some user-centered design and Usability challenges (presented as Wicked Problems) while designing for interactive exhibitions are that installations have to be intuitive and self-explanatory, while at the same time having multiple target groups with varied knowledge and needs [32]. The term *immediate apprehension* is therefore an important aspect of exhibition design [34].

3.3 Designing for Large Language Models

Large Language Models (LLM) are Artificial Intelligence (AI) systems that generate coherent and context-based strings of text following grammar, syntax and semantics of different languages [5]. LLMs are fed large amounts of training data to inform their responses.

There are several factors regarding the trust and quality of interaction of LLM behavior that affect the user experience [3]. One such factor is human-like features, anthropomorphism, in AI. Contradicting research suggest that users in some cases prefer high anthropomorphism [3], while other suggest that text-based chat-bots are perceived as more trustworthy [37].

Another challenge of AI interactions concerns AI-bias, where biased training data can lead to discriminatory outputs and have a negative effect on society [4], [5]. Moreover, some LLMs have a tendency to produce something referred to as "hallucinations" [5], which is when an LLM produces false or misleading information that appears correct. Hallucinations can amplify biased training data, further spreading misinformation [5].

Generative AI excels at repeating ideas of old [38], and through the lens of *Hauntology*, these amplifications of biased training data and average outputs can influence design practices [39], [40]. The term Hauntology comes from Derrida’s philosophy [41], which was further explored by Fisher [42], and it is the phenomenon that ideas of the past haunts us like ghosts, shaping the present. Heron et al. [39] argues that there are three main causes that makes AI fundamentally hauntological: it is probabilistic, generating outcomes through matching patterns; it is a mirror, reflecting our own culture and biases; and it converges towards reasonableness, producing uncontroversial and widely accepted outputs. The hauntological perspective on generative AI describes the notion that the continuous usage of AI in design practices will only reinforce old ideas instead of allow exploration of new ones [39], [40]. However, Patil et al. [40] mean that the metaphor of Hauntology within Speculative Design can be a helpful tool in dealing with imaginary, future solutions and uncertainties.

Furthermore, state of the art AI-models that are used by companies or individuals for *daily tasks*, *efficiency*, or *emotional support*, show concerning issues in being highly sycophantic, based on a study by Cheng et al. [6]. The authors state that AI models are shown to affirm the users beliefs 50% more than humans would, even on topics such as manipulation or relational harm. In the same study, participants preferred the sycophantic AI model; they favored its responses, trusted it more, and were more likely to use it again.

One of the many challenges with AI-ethics and what makes the de-mystification of AI into a Wicked problem is the concept of the *Black box*. It refers to a system or technical solution where the user is only aware of the in- and outputs [43], lacking understanding in the internal processes. Lee [44] suggests that algorithms are powerful black boxes. Arguing that they can be understood both as fluid and assembled, or as a defined system that can be made transparent and held accountable.

3.4 Ethical and Social Considerations

Ethics is the philosophy of morality that explores the rights and wrongs, as well as key principles that dictate human conduct [45]. In terms of AI-ethics, a report by the European Commission [7], followed by a revised self-assessment list for trustworthy AIs (ALTAI) [8], was created as a way to self-assess and thus ensure ethical creation of AI models.

These guidelines presented by the European Commission [7], [8] are: *human agency and oversight*, which means that AI should strengthen and assist humans in making decisions and empower human agency; *technical robustness and safety*, which means that AIs requires dependability and resilience to ensure the system behaves reliably, as intended, and safely under changing conditions; *privacy and data governance* which means that the creation and use of AIs should not infringe on peoples privacy and personal data; *transparency*, which means that the AIs development, system and training data should be fully transparent and publicly available; *diversity, non-discrimination and fairness*, which highlights the importance of eliminating unfair bias, and fostering diversity in all aspects of its use and development; *environmental*

and *societal well-being*, which focuses on the sustainable development of AIs for the environment and thus future generations; and lastly, *accountability* that advocates the enforcing of accountability for the AI systems and its outcomes.

Although this is a critical design project that aims to generate ethical insight through reflections on human-AI interaction, any results regarding ethical insight can be upended and instead be used to design for trust in a system that does not provide ethical, human-oriented design. Frauenberger et al. [46] stresses that ethics must be the underlying perspective for research projects.

As further described by Frauenberger et al. [46], there are several ethical challenges in HCI research, some of which will be highly relevant throughout this thesis work: *Exploitative Research* creates unpredictable processes and outcomes which risks unperceived ethical dilemmas; similarly in the *Context* in which the research is conducted; *Protecting Participants*, as informed consent is not always adequate during uncertain work; *Values and Provocation*, as the intersection of critical design, HCI and art has shown that projects do not always embrace ethical values and instead provoke and critique; *Stakeholder ethics*, as interdisciplinary work might include different ethical positions; *Outcomes and Expectations* might differ, as various stakeholders measure success differently, there is a challenge in ethically managing expectations; The researchers role as an objective bystanders is a debunked notion, therefore, the researchers *Closeness*, role and power position needs to be considered; finally, *Exits*, whereas this is only a slight risk in this project, but if the project report creates surge waves of interest after the project is done, this might create new ethical dilemmas.

3.5 Designing for Uncertainty

The last but the perhaps most relevant conceptual framework is how to design for uncertainty. This thesis work was framed by it, and required an open and creative approach to methodology, as well as iterative development. Uncertainty in design practices can have many forms according to Daalhuizen et al. [47]. Categorizing the causes as attributed to the: *individual*, *social context* or *task* itself. As well as the approaches to uncertainty being: involving *stakeholders*, *keep going*, *articulating business values*, *framing the problem*, *create open communication culture*, *use of intuition* and *working multidisciplinary*. This is described by Daalhuizen et al. in figure 3.1, [47], where the causes are described as the context surrounding the non-routine situation and the approaches as knowledge-based behavior.

Uncertainty in design practices can be seen as an asset in RtD [48], [49]. By employing an oscillating approach, going back and forth between different uncertainties, the designer creates anchors of clarity that drives the design forward. This oscillating approach is an effective way of dealing with uncertainties [48]. Epp et al. [48] exemplify this with rigorous user research as an anchor to allow for an uncertain ideation in speculative design work. Oscillating between uncertainties can also be a method of finding previously unperceived uncertainties [49]. This phenomena is visualized by Epp et al. in figure 3.2, [48].

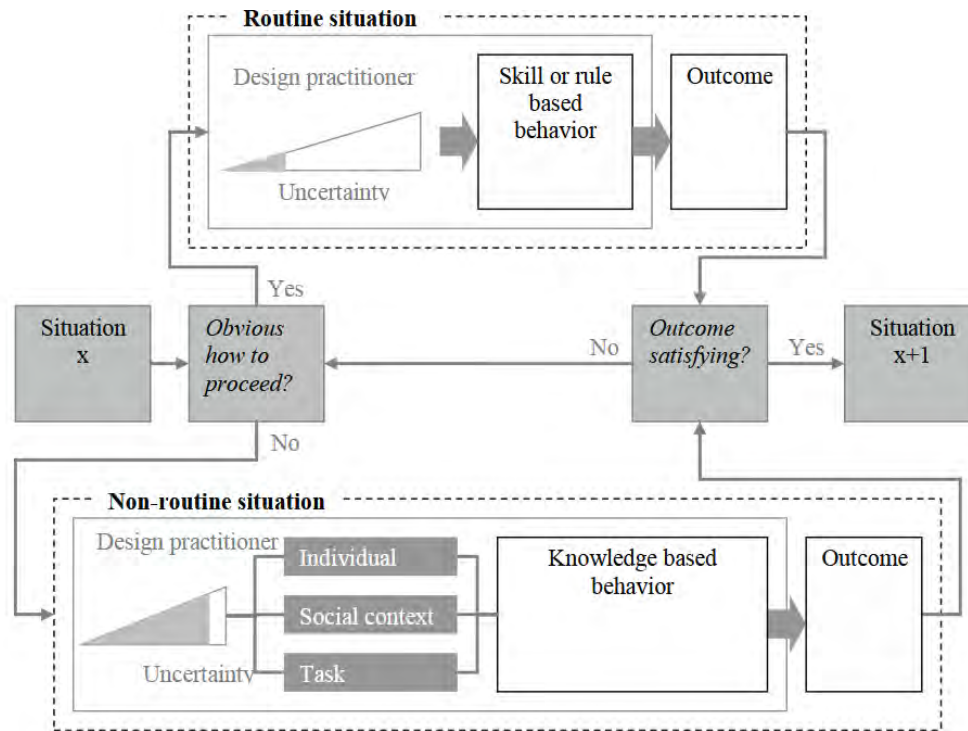


Figure 3.1: Dealing with non-routine design problems, by Daalhuizen et al. [47]



Figure 3.2: Uncertainties as Generative Resources in RtD, by Epp et. al [48]

Furthermore, through focusing on non-immediate perceived uncertainties the researcher can find alternatives to predictable design process paths and generate new knowledge and possibilities for design [48]. New design spaces emerge from uncertain RtD, and these spaces stimulate creativity as well as more resilient and adaptive design practices [49].

Ultimately, uncertainties within the artifact itself can also serve as an advantage, enabling the creation of artifacts that invite deeper user reflection. Gaver et al. [50] mentions the concept of ambiguity in HCI. The authors state that by embedding ambiguity into an artifact, designers can generate interpersonal relationships between the user and the artifact, ultimately provoking critical reflection through the interaction itself. By leaving things out, people are invited to fill in the blanks. The invitation to form own ideas, in turn creates a personal connection to the piece.

4

Methodology

In this chapter, methods that were found to be suitable for designing for the unknown are described. Furthermore, this chapter contextualizes each method, explains why it was used, and discusses its benefits. Thereafter, the next chapter, 5. Results, will present the results gained from the methods and each step in development. In the beginning, each stakeholder and team member had different views on what the project entailed, what was going to happen, and where it would end up. The process required an openness among the team, an openness to discover and learn, but also to fail. Uncertainty was the driving force.

The methodology in this project was developed incrementally, as oscillating between different possible uncertainties and design ideas can reveal new possible directions [48], [47], see figure 4.1. This illustration of working with uncertainties is inspired by Epp et al. [48]. It describes the starting position (star node) and the immediately visible uncertain design possibilities (black node), these were navigated during the initial step of covering stakeholder views. They are uncertain in the sense that what comes after is not clear until they have been explored.

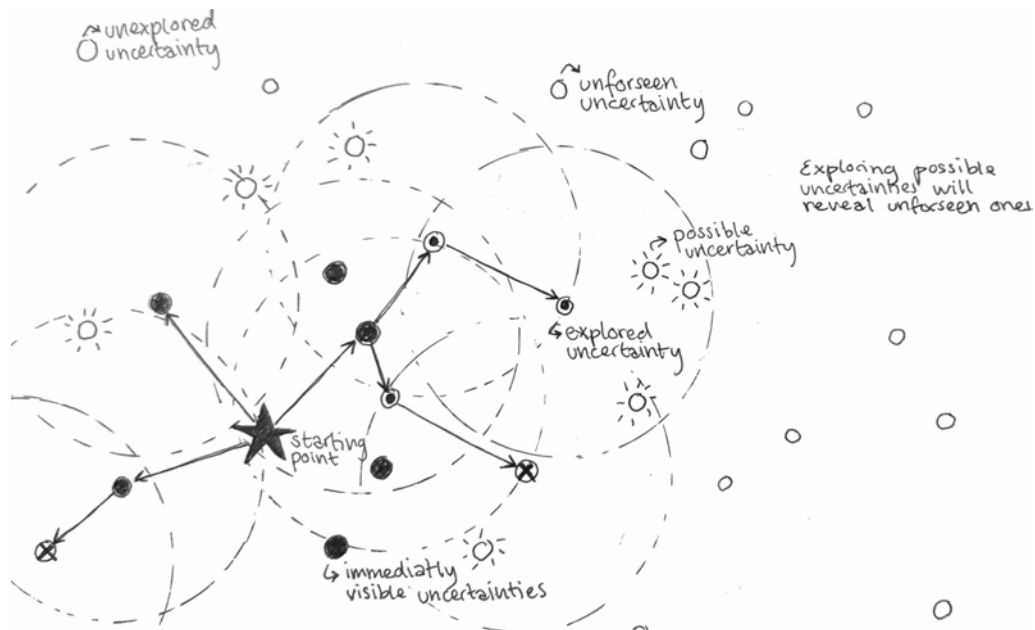


Figure 4.1: Illustrating how exploring uncertainties can reveal new ones, creating an unpredictable path and process

Exploring development in different directions reveals new design possibilities as the scope of vision expands. Many dead ends and failures (X node) might be met, but all the uncertainties explored reveal new possible outcomes. Only trial and error can expose those unforeseen uncertainties, and some might never be explored due to what path is chosen. This is an illustration that attempts to exemplify how this project dealt with an ever changing scope, solution space, and design possibilities through constantly iterating and embracing the unknown.

Each iteration can be generally described as a single loop. The parameters of the previous iteration defined the input for the next, laying the foundation for the exploration of the design space and its uncertainties. The point of convergence was reached through engagement with stakeholders and users, which then defined the parameters for the next iteration, see figure 4.2. Iterative and collaborative work is part of *Agile development*, an approach to software development that is not conforming and procedural [51]. The authors define success of an Agile development project as whether or not you have delivered something of value, rather than if you have followed a predefined procedure. They stress that Agile teams have agency over their own processes, instead of sticking to pre-defined methodology.

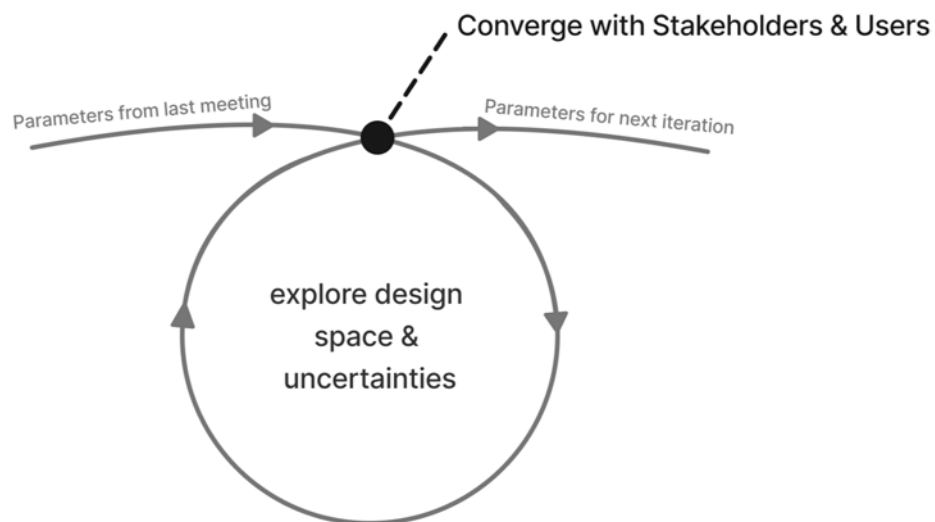


Figure 4.2: Overview of the iterative approach-loop

The methodology and process, as well as the different methods utilized in the project will be described in the following sections, see figure 4.3 for a visual overview of the process. The first iteration, visualized in red, started with planning and converging views: defining value oriented goals, understanding user and stakeholders, as well as defining the experience. This iteration also contained ideation and concept creation. Later, returning to converging views; receiving stakeholder feedback and aligning project goals once more.

The second iteration, visualized in blue, continued development of the chosen concepts and visual elements, introducing user testing in controlled settings and reporting back the findings to stakeholders, again using this as an opportunity to further define the project scope, purpose and direction.

The third iteration, visualized in yellow, introduced front-end development and connecting it to the back-end server, as well as continued development of visual art assets. This iteration included formative on-site audience testing in the exhibition space *Wisdome*, at the science museum *Universeum*. The user feedback from the tests combined with stakeholder co-creation sessions informed the last iteration.

The fourth and last iteration, visualized in purple, now had clear design goals and was not exploring any new uncertainties, but rather focused on making everything work the way intended. This sprint included a smaller on-site audience test and ended with a final showcase and collection of summative user feedback.

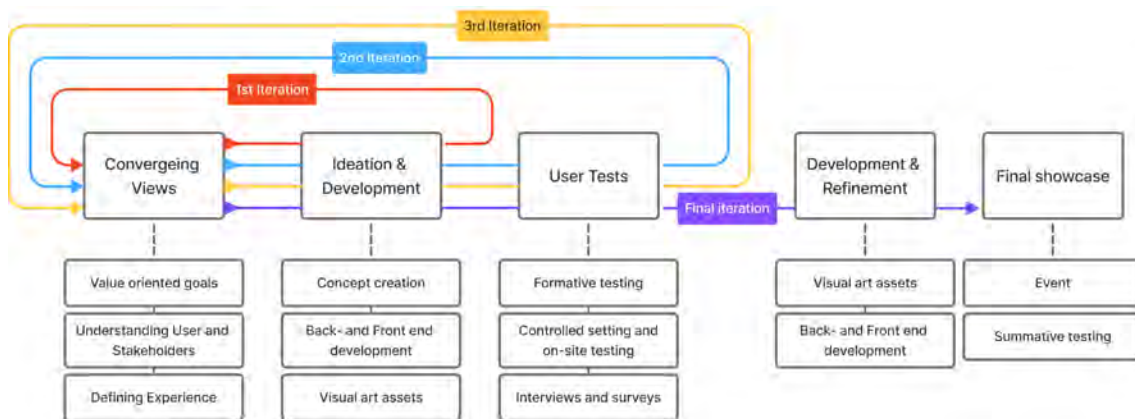


Figure 4.3: Visual process describing all design iterations

4.1 Converging Views

Considering the many stakeholders in the project, a joint understanding of the desired effect, and what elements would create it was needed. *Mind-mapping*, a non-linear method for generating ideas by visually mapping the relationships between concepts within a specific topic, thus creating a system of connecting ideas [52], was used as a tool to structure important elements for the desired experience, see figure 4.4. This method was good for gaining a better understanding about a specific topic and how different themes are connected to each other, as it helps the early exploratory phase of design projects [53].

Thereafter *co-creation* sessions were held to focus on shared values, presenting and converging everyone's ideas. Co-creation is described in a general sense by Sanders and Stappers [54] as "collective creativity, i.e. creativity that is shared by two or more people." The authors also mention that as a sub-category of co-creation is co-design, which is described as the collaboration between designers and non-designers in any part of the design-process. When working on complex projects, having a multidisciplinary team utilizing collective creativity helps to gain a more complete understanding of the problem [47, p. 153]. The shared value-oriented goals defined after the co-creation session were differentiated as *Magic*, *Wow*, and *Effect*. They were used as central anchor points between all stakeholders to align the artistic intention throughout the technical development. They will be thoroughly described in chapter section 5.1.1.

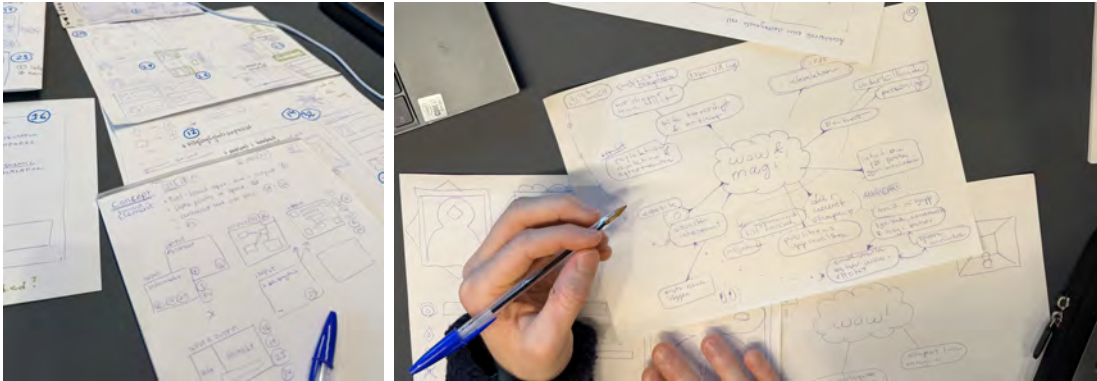


Figure 4.4: Mindmapping value-oriented goals and ideation process

Together with Scionova and the project lead, a *User Story Mapping* session was held to define functions for the backend through defining users needs and the desired experience. User Story Mapping is a visual method for Agile development, to help define what work is needed to create value in the user experience of a product or service [55]. The authors describe the story map itself as made up of two axes; on the x-axis is the narrative flow, on the y-axis are all the user tasks required to use the product or system. By mapping the user story, an understanding of both the requirements needed for the user interaction and a collective understanding of the user flow was gained, which aligns the views throughout the team [55].

Following this, two *User Journeys* based on two different scenarios for the showcase were constructed. These scenarios were *Scenario 1: Installation*, and *Scenario 2: Wisdome*, further described in section 5.1.3. Similar to User Story Mapping, User Journey mapping is a design method intended for the early stages of a project to reveal relevant needs and requirements for the system and its users by mapping the processes in which they participate [56]. By making User Journeys in the early stages it was possible to effectively plan ahead before entering the next phase.

Several interviews with the artists MF & WS were held to get an in-depth understanding of the artistic intention. This helped define the scope and make an initial visual description of the system as a *human-machine system*, inspired by the ACD³ framework by Bligård [57]. A human-machine system is described as a set of humans and machines interacting with each other, within a defined environment, in which they perform tasks to reach a set goal [57]. The author stresses that to fully understand a complex system, you have to study its entirety, meaning the user, machine, environment, and how they interplay, to gain a complete understanding of its effect and ability to achieve certain goals. The system image also helped to define the design deliverables as well as the design criteria.

Before concept ideation began, *Moodboarding*, a method of gathering images and abstract media to assemble for inspiration in the design process [58], was used as a tool to align our views, as well as the views of the artists. This technique is thought to facilitate a creative dialogue, and sparks inspiration between the designers and the clients [58]. The initial stages also included a visit to Wisdome, to understand the space and receive tech-specifications, see figure 4.5.



Figure 4.5: First Wisdome visit

4.2 Ideation

The next phase consisted of ideation, low-fidelity prototyping and early evaluations. Two *Brainstorming* sessions for *Scenario 1: Installation* was carried out. Brainstorming is an ideation technique in which a team develops ideas or executes a specific task within a problem context, by participating in unrestricted discussions [31]. Each Brainstorming session consisted of sketching an idea for five minutes. This was repeated three times. The first session focused on the whole problem context, and the second was constrained to the UI, see figure 4.4. The goal of Brainstorming is to generate as many ideas as possible, increasing the probability of coming up with one or more quality concepts [31]. After each session, the ideas were discussed, identifying the pros and cons of individual solution elements. Each element was assigned to a function in a matrix, and the most fundamental and innovative highlighted.

These were combined into four initial concepts, and were chosen to reflect at least one *safe* option, one *slightly experimental* and one *highly experimental*. The functions were picked to include a wide range of solutions in order to evaluate a broad spectrum of design. Wireframe sketches were used to get a rough idea of the concept and its intended navigation. *Wireframing* is a method of sketching out the foundation of an interface, mapping out important topics like layout, structure, and user flow as described by Hartson and Pyla [31]. Wireframes are usually done with simple shapes and text, either with pen and paper or using digital design tools. The authors argue that wireframes are effective because they quickly grant an overview of the interface and the intended interaction, without wasting tools or resources.

Thereafter similar ideation sessions were held for *Scenario 2: Wisdome*. The concepts from both scenarios were narrowed down to two each, and an additional two concepts focusing purely on interaction was created to support *Scenario 2: Wisdome*. All six concepts were refined and visualized using digital prototyping tool *Figma*. *Prototyping* is a design method which rapidly transforms ideas into tangible drafts, to efficiently test and tweak a concept [31]. The author describes a prototype as a representation of a design, which come in different fidelities. These initial concepts were of *low-fidelity*, commonly crafted in early stages of projects, to get a realistic, yet quick, overview of solutions. Low-fidelity prototypes were effective in communicating and comparing design ideas to stakeholders during in-house evaluations.

A promotional video was created evaluating the interaction flow, testing an early prototype for AI generated voices synced with the generated text. The video also included a later iteration of the visual description of the system to better communicate the system architecture.

4.3 Early Formative User Tests

In a meeting with all stakeholders, the six concepts were presented and discussed. Based on co-creation sessions, *Scenario 2: Wisdom* was chosen, and the concepts were refined into three, focusing mainly on the UI. From early formative user tests, and continuous co-creation sessions, a final concept was decided on. A *high-fidelity prototype* was created to be able to test the look and feel, details, and interactivity, before the final implementation.

Thereafter two in-depth *group interviews* were held, Performed in a controlled setting to collect relevant user data. *Interviews* are a data collection method where you talk with a person of interest about a specific topic, with different levels of pre-determined structure, to collect information about said topic [17]. Probing was used to ensure the interviewer reaches an in-depth understanding of the user’s perspective. The interviews had the roles interviewer (facilitator), note-taker (observer) and participants (users). The observer transcribed the interview live to create a foundation for future data analysis. The participants were chosen to as expert reviewers with knowledge in usability, software development and cognitive science.

Group interviews were chosen as a method to facilitate discussion among participants [17]. The facilitator had an agenda, but the general structure was flexible to follow unanticipated discussions as they were brought up. This was an effective method to understand collaborative interactive experiences and bring up perspectives from different user stories. The interviews were of *semi-structured* nature, utilizing both pre-determined questions and open-ended discussions to gain both specific data and rich insights about a set topic [17]. Semi-structured interviews ensures that the participants stay on topic while still allowing them to shape the direction of the interview.

The aim of this phase was to collect qualitative data to inform future design decisions, see appendix A.5 for the interview questions. These were formative tests that identified general problems in the interaction and user experience.

In the user tests, the scenario was simulated through scripted dialogue. The gathered data was analyzed using *Thematic Analysis*, a method for analyzing *qualitative data*, and recognizing patterns in a dataset [59]. Following the framework created by [60], the dataset was efficiently broken down large into smaller sections, in which overarching and recurring themes were analyzed and mapped in order to better understand the collected data. The data gathered mostly focused on UX relevant themes, but also included themes relevant for back-end development. These results established a foundation for the upcoming concept development.

Data collection can be *qualitative* or *quantitative*, with the purpose of being *formative* or *summative* [17]. Qualitative data cannot be measured, while quantitative data are metrics that can be measured and compared. Interviews generally only collect qualitative data, while surveys are more effective for collecting quantitative data. Formative evaluations is when data is collected with the intention to inform future design work, summative evaluations are carried out on a finished product to assess success. Collected data needs to be analyzed in order to generate meaning.

4.4 Concept Development

The concept development phase consisted of three main areas: visual components, front-end, and supporting back-end development. This phase also included an internal evaluation at Wisdome without an audience.

Two visual components was created in dialogue with MF & WS, namely portraits and background visuals. Exploring different styles and techniques through several iterations the art assets were evaluated and tested for scale, readability as well as artistic intent in the Wisdome. The portraits were sketched by hand and modified using commercially available online ASCII editors and *Adobe Photoshop*. The background visuals were programmed using a open-source JavaScript library visual code editor called *p5.js*, and edited to match the fulldome-format using *Adobe After Effects*, which was then evaluated using *Amateras Dome Player* program.

The front-end was coded using the open-source front-end JavaScript library *ReactJS*. This allowed for seamless implementation of user interface elements. Initially a mock-server with pre-generated responses was used to evaluate the UI elements. Thereafter, the server connecting the LLMs to the UI was implemented, which allowed real-time testing. The more further along development the prototype was, the more issues became apparent and could be addressed. The code was written with the help of AI-code agents. To support the back-end development, several meetings with all stakeholders were held, see figure 4.6.



Figure 4.6: Stakeholder meetings and consultation

4.5 Audience Tests and Iterative Refinement

The final step of development consisted of two on-site audience tests, surveys, and defining design goals as well as necessary functions based on feedback. This was crucial to ground the development through feedback from real experiences.

A large-scale user test in a realistic on-site scenario was held to evaluate the prototype's current stage. Two different background visuals were tested as well as four different alternatives for the portraits, three of which had corresponding GIF versions. Among the audience were employees of Universeum, students, faculty members, family and friends. Around 40 people attended the user test, which collected 24 responses in a Survey filled in after the test. The prototype and test results and analysis will be presented in section 5.4. Audience Evaluations.

Surveys, or questionnaires, are a method of gathering information, through closed or open-ended questions, about a topic of research [17]. As the survey could be easily distributed it could therefore gather a larger amount of data than through interviews. There is however no facilitator who can provide explanations, and therefore the formulations need to be carefully constructed, see appendix A.6 for the survey questions. Gathering the quantitative data from surveys was done through a number of ways. To be able to compare a range of user responses, *Likert* scales were used [17]. The survey also collected metrics such as: "Describe your experience in three words". There were also free text fields for some open-ended questions. The survey consisted of 14 questions and was designed to generate feedback on overall experience, believability, visual elements, as well as demographic questions focused on AI-usage.

The user test was followed by analyzing the audience feedback and stakeholder feedback through Thematic Analysis. The groupings were then reviewed again, focusing on the corresponding functions needed to be implemented. This laid the groundwork for future development. The identified problem areas largely focused on creating a more dynamic conversation, user interaction, and having more variation in visual effects.

An updated User Journey focusing on event-flow and required system functions was constructed. A meeting was held together with the backend engineer master student to coordinate efforts and set clear design-goals and functions that needed implementation. These functions were weighted based on the *MoSCoW* framework, a prioritization method in which you align with stakeholders on the projects requirements [61]. Dividing the functions between *M-must have*s, *S-should have*s and *C-could have*s. This was done to ensure there was a consensus in the team on which features were to be focused on, within the given time frame. *W-wont have*s were not denoted in this project.

Through design iterations, the front-end and visual components were further iterated on based on received feedback from users. This, in combination with collaborating with musician Jakob Dahl to generate a background music track, were efforts made to foster greater immersion and aesthetics. Apart from layout changes to the front-end, a separate front-end was coded together with Tengman to give the on-site developers in the tech booth more control.

A second, smaller on-side audience test was held at Wisdome, Universeum approximately two weeks after the first test. The second test focused on evaluating the newly implemented functions and overall flow of the human-to-AI, as well as the curator-to-audience interactions. A survey generating qualitative answers was conducted, since quantitative answers would not be statistically viable with few (2) audience members, see appendix A.7 for survey questions. Analyzing the feedback, another list of design goals was constructed by the front-end and back-end team.

4.6 Presentation of finished Design

In preparation for the final showcase, two mock tests of the showcase was performed in a controlled setting, outside of the exhibition space, to ensure a seamless audience experience by identifying any potentially overlooked technical or facilitation issues. The mock-test was performed by enacting several likely to extreme scenarios. The second mock-test was performed in a similar manner, after completing bug-fixes identified in the first mock-test, thus completing final refinements.

The final showcase was held in Wisdome, Universeum, in front of a live audience of 57 people, where the survey collected 39 responses. The event was carried out with one human curator in charge of the human-to-AI interaction as well as facilitation of the audience interaction with the system. Two team members were in the control room, manually forcing state changes when necessary. The controls were also available to the curator through a small portable controller. The control room operators were also in charge of setting up and controlling the background visuals, music, volume and lighting. The resulting showcase and audience feedback will be described in further detail in section 5.6 Final Showcase.

The showcase ended with a survey to collect summative user feedback evaluating the system, interaction and experience. The survey was largely structured similarly to the one from the first user test to be able to contrast and compare answers. It consisted of 12 questions focusing on collecting quantitative data, though there were also free text fields available for open-ended questions, see appendix A.8 for survey questions. The demographic questions focused on AI-usage, and attitude towards generative AI in general. The survey was analyzed using Thematic Analysis, grouping the answers based on different elements which affected the experience and general audience reflections.

5

Results

This chapter presents the findings gathered through the iterative research process described in the previous chapter. The insights combine both qualitative and quantitative user data, as well as design decisions based on said data and stakeholder meetings. The results chapter largely follows the structure of the methodology chapter, and contains the following sections: *5.1 Defining the Project Scope*, *5.2 Initial Concepts*, *5.3 Prototype Implementation and Development*, *5.4 Audience Evaluations*, *5.5 Final System Design & Functions* and *5.5 Final Showcase*.

5.1 Defining the Project Scope presents the results of the methods described in section 4.1 Converging Views; the joint value goals, User Story Mapping and User Journey for the different scenarios. The scenarios were explored through ideation and stakeholder meetings, as will be described in *5.2 Initial Concepts*, that presents the results of section 4.2 Ideation & 4.3 Early Formative User Tests. The initial concepts were evaluated in group interviews whose feedback informed the following section *5.3 Prototype Implementation and Development*. Here we present the results of the methodology described in section 4.4 Concept Development, where the front-end development and the implementation of a high-fidelity prototype, as well as art assets development and supporting back-end development are described. The results of the methodology section 4.5 Audience Tests and Iterative Refinement is presented in *5.4 Audience Evaluations* and *5.5 Final System Design & Functions*, describing the two on-site evaluations of the high fidelity prototype that informed the final system design. Finally, *5.5 Final Showcase* presenting the results of section 4.6 Presentation of finished Design, describes the final system and the resulting summative user test.

5.1 Defining the Project Scope

To deliver an immersive experience, believable conversation and engaging interaction, several methods, as described in chapter 4. Methodology, 4.1 Converging views, were needed. This section presents the findings from the methods used to align the team's views, which defined the project parameters in its early stages.

5.1.1 Magic, Wow and Effect

A joint understanding focused on trust in the system, reflection on content, and a critical view of AI resulted in three value oriented goals: *Magic*, *Wow* and *Effect*. They were defined as follows:

Magic: to have a realistic, believable conversation. Magic is what is generated in the interaction between the user and the system. To spark magic in the interaction, the user needs to trust the system and its content. Therefore, the interface needs to support the user in both being believable in its content, as well as being aesthetically interesting.

Wow: created through the audience interaction and atmosphere. Wow is the reaction to the interaction between the user and the interface, meaning the reaction to the Magic. The atmosphere should be eerie; balancing between hopefulness and hopelessness, and provide contrasts between machine "cold logic" and humanities "warm intelligence".

Effect: as critical reflections on AI. The effect is the reflection the user has when leaving the interaction with the system. This is how the user shares the experience with others. This aims to provoke critical thought about possibilities and risks of technology, trust in AI, and general reflections about the dialog.

5.1.2 Story Mapping and System Description

The collaborative user story mapping session held with the project lead and Scionova led to defining the initial parameters and flow of the interaction, see figure 5.1.

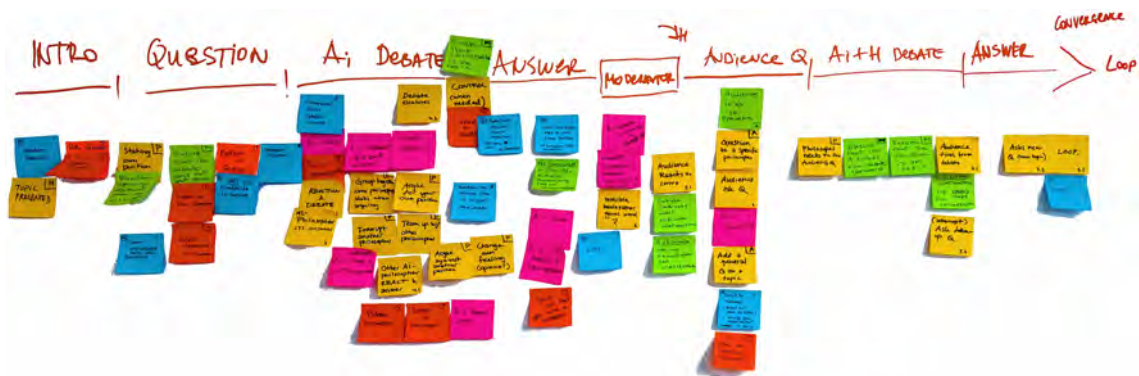


Figure 5.1: Story mapping session with post-its

The results of the story mapping session were further iterated on. The post-its were translated to the digital world and color coded to map relevance and function. The commentator and moderator role was not yet decided to be either human or AI, and if it was going to be AI, if it would be visible or invisible in the system.

The flow of interaction was defined as: Event Start; Introductions; Opening Question; AI Debate, Answer & Moderator Intervention; Audience Question; AI & User

Debate; Loop back to Audience Question; Event End. The story mapping session was conducted on the basis that there was going to be a curator to facilitate the event. Based on this information, The human-machine-system was created, see figure 5.2. The system functioned as follows:

1. The audience and curator provides: questions/prompts/themes and can: interrupt/change topic if necessary.
2. The visible or invisible moderator/commentator moderates the question to one, or several AI-philosophers, based on their knowledge domains.
3. A conversation/debate in between the AI-philosophers is had.
4. When the conversation/debate reaches a conclusion, gets out of hand or starts to loop, the curator interrupts with a new stored prompt from the audience (if available) or comes up with its own theme, based on pre-written prompts.
5. The output is the take-aways from the conversations between the AI-philosophers.

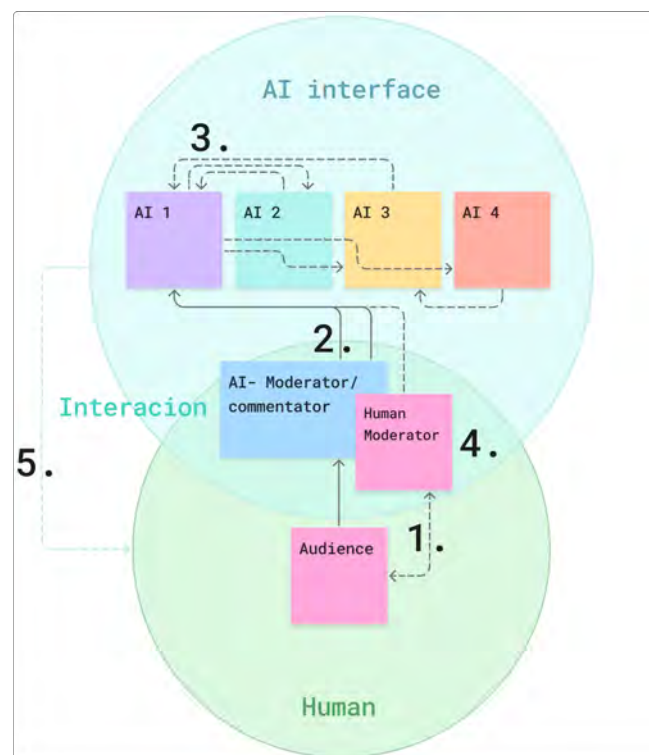


Figure 5.2: Initial human-machine-system

There were however two potential solutions for the event structure being in a curated event or ongoing museum installation. In order to further evaluate the different scenarios, two user journeys were constructed.

5.1.3 User Journeys

Two user journeys were constructed in order to decide on the next step forward in defining important aspects of the interaction, and to start discussing feasibility of technical parameters, such as the back-end system and possible front-end interaction. *Scenario 1: Installation* describes an ongoing museum installation where the user comes and goes into an already ongoing active system. *Scenario 2: Wisdome* describes a curated event in which the user arrives at a set time and stays for the duration of the show.

Scenario 1: Installation's user journey is visualized in figure 5.3, and the flow is described underneath the figure in step 1-6.

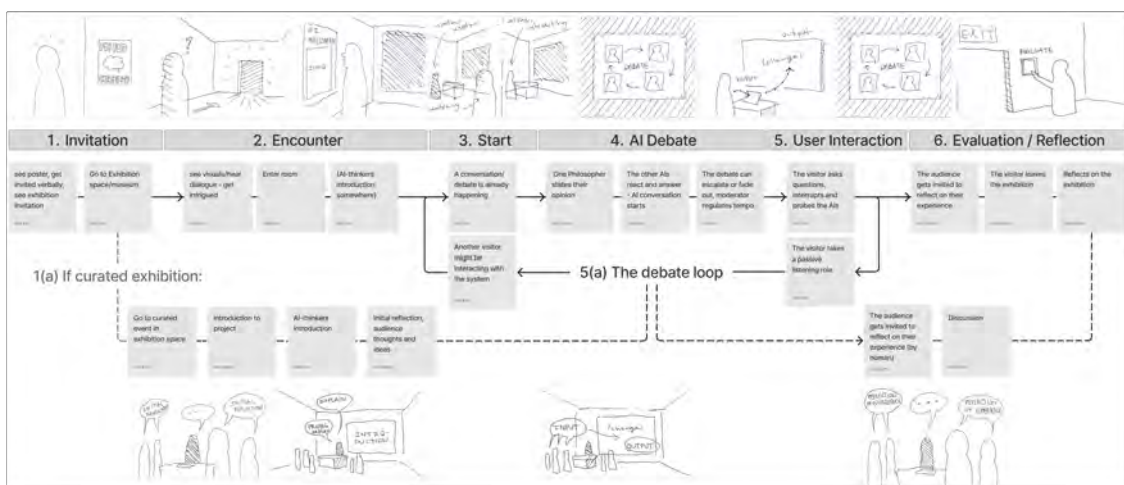


Figure 5.3: User Journey of *Scenario 1: Installation*

1. **Invitation:** The visitor gets invited to the installation: sees poster, is verbally invited or sees showcase invitation and then goes to the exhibition space.
 - (a) **If curated exhibition (e.g., a public showing or guided tour):** The visitor goes to the set exhibition time and gets introduced to the project and AI-philosophers. Initial reflection is had with the audience. The curator then controls the interaction with the AI-philosophers, go to **5(a). The debate loop**. The curated event ends with audience reflections and discussions.
2. **Encounter:** The visitor gets intrigued to enter installation space through seeing visuals/hearing dialogue. Entering the room the AI-philosophers introductions are available as a part of the interface or physical installation.
3. **Start:** A conversation/debate is already ongoing, another visitor might be interacting with the system.
4. **AI Debate:** The debate is constantly ongoing and never pauses, the AI-philosopher reacts to each others opinions and answers. The debate can escalate or fade out. The curator regulates tempo and topics if there is no visitor interaction.

5. **User Interaction:** The visitor can ask questions, interrupt and probe the AIs, alternating between a participating and passive listening role.
 - (a) **The debate Loop:** Each question starts a new debate and loops back to 3. *Start*. This loop is endless.
6. **Evaluation / Reflection:** When the visitor is ready to leave they are invited to reflect on their experience as a part of the physical installation.

Scenario 2: *Wisdomes'* user journey is visualized in figure 5.4, and the flow is described underneath the figure following step 1-6.

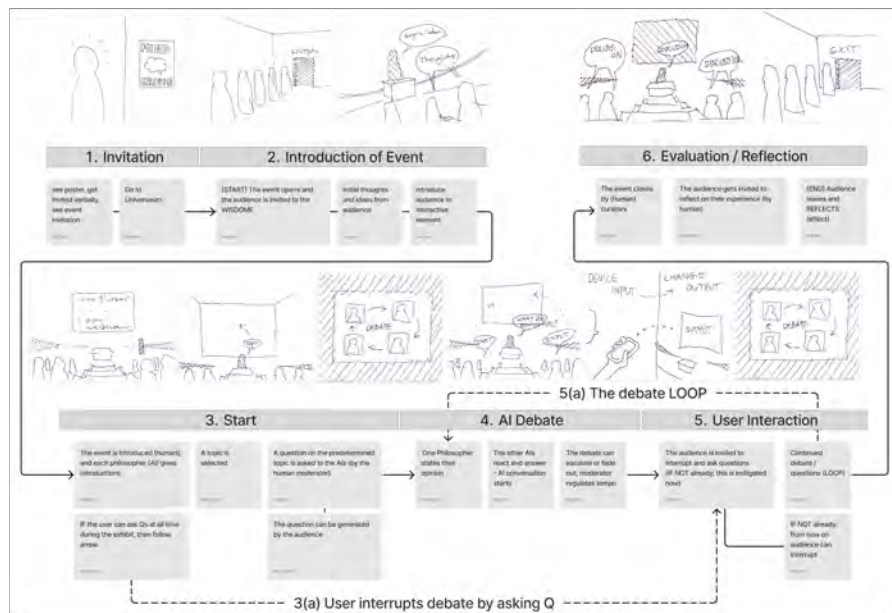


Figure 5.4: User Journey of *Scenario 2: Wisdomes*

1. **Invitation:** The user gets invited to Wisdomes: sees poster, is verbally invited or sees event invitation. They go to Universeum.
2. **Introduction of Event:** The event opens and the audience enters to the Wisdomes. Initial thoughts and ideas are discussed with audience and they are introduced to the interactive element.
3. **Start:** The event is introduced by curators, and each AI-philosopher gives their introductions. A topic is selected and a question on the predetermined topic is asked to the AIs by the curator - these questions can be generated by the audience.
 - (a) **User interrupts:** If the user can ask questions at any time during the exhibit, then skip to the debate-loop during this part.
4. **AI Debate:** The debate starts as one AI-philosopher states their opinion. The other AIs react and answer, and the conversation starts. The debate can escalate or fade out. The moderator regulates the tempo.

5. **User Interaction:** The audience is invited to interrupt and ask questions. If not already: this is initiated now, and from now on the audience can interrupt at any time.
 - (a) **The debate Loop:** Each question starts a new debate and loops back to 4. **AI Debate.** This loop continues until event closes.
6. **Evaluation / Reflection:** The event is ended by curators. The audience is invited to reflect on their experience and leaves Wisdome.

5.2 Ideation & Initial Concepts

The ideation was based on design criteria from the stakeholders, see appendix A.1. The ideation resulted different solution elements mapped to product functions, see appendix A.2, that were combined into six initial concepts. Four concepts addressed *Scenario 2: Wisdome*. They were divided into visual concepts, which encompasses the background dome visuals and live-feed screen, and interaction concepts addressing possible mobile app solutions. The remaining two concepts focused on *Scenario 1: Installation*. See appendix A.4 for a description of the initial concepts for *Scenario 1: Installation*, and A.3 for initial concepts for *Scenario 2: Wisdome*. Based on the results of the co-creation sessions for the six concepts, key design decisions were made to advance the process.

5.2.1 Design Decisions

Firstly, the team decided to move forward with *Scenario 2: Wisdome*. It clearly defined the scope of the back-end prototype, and was seen an opportunity of working in a new and unexplored format tied into the already uncertain outcome of the project. Secondly, the interaction concepts were scrapped, as mobile apps could take away from the intended immersion. The team instead decided that using a microphone and voice input for the user interaction as it would simulate a conversation by talking directly to the system.

Lastly, from the visual concepts the team agreed on using portraits on the output-screen, since they gave a clear understanding as to who was speaking and would map the portraits and names together through color-coding, as in *Visual Concept 2*, see figure 5.6. The team liked the code-aesthetic as it gave a clear reference to the topic of LLMs, a retro-feel as a homage to Home of the Brain, and also showcased transparency in that it was AI-models and not real philosophers discussing as in *Visual Concept 1*, see figure 5.5. Regarding background visuals, the team enjoyed both minimalistic and eccentric visuals, and needed more refined concepts to make clear decisions. However, using abstract representations for the AI-philosophers as in *Visual Concept 1* was preferred for the sake of immersion.

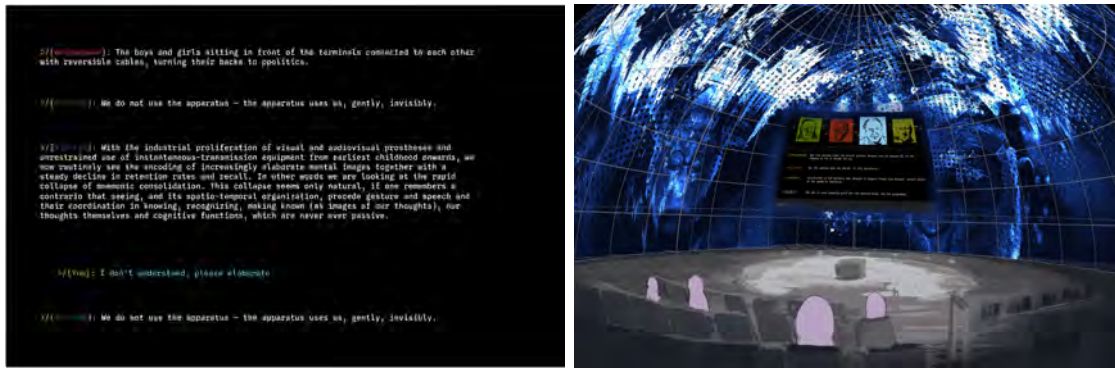


Figure 5.5: Visual concept 1: text-based UI, and background visuals in dome-setting

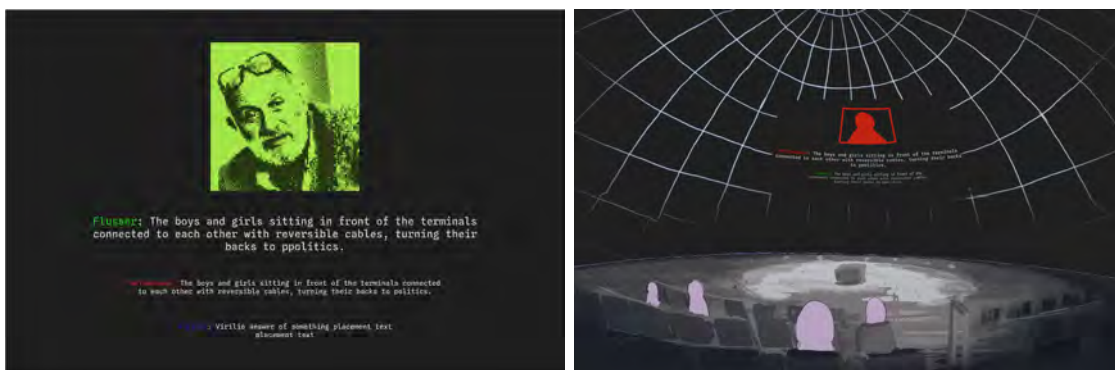


Figure 5.6: Visual concept 2: text-based UI, and background visuals in a dome-setting

5.2.2 Concept Iterations

Merging the two concepts in the previous section resulted in three refined concepts, each focusing on the live-feed output-screen, see figure 5.7, 5.8 and 5.9. These concepts were discussed and prioritized in stakeholder meetings, to be able to test the interface design with users.

The team preferred having all four AI-philosophers portraits on screen at the same time to create the feeling of a conversation between all, as well as help users distinguish between them, as in figure 5.7. However, there was a preference for the ASCII style portraits as in figures 5.8 and 5.9. Keeping the color-coding of the AI-philosopher names combined with the portraits was of high importance to clearly distinguish between them and to understand who is currently speaking.

Several other features were discussed which needed validation through user testing, specifically from a cognitive ergonomics and visual hierarchy perspective. Some of the features discussed were: active AI-philosopher text placement, text movement top-to-bottom or bottom-to-top and helpfulness of conversation history. Based on the team discussions, the concept presented in figure 5.7 was prioritized in the user testing, and the concepts as seen in figures 5.8 and 5.9 were shown as alternatives.

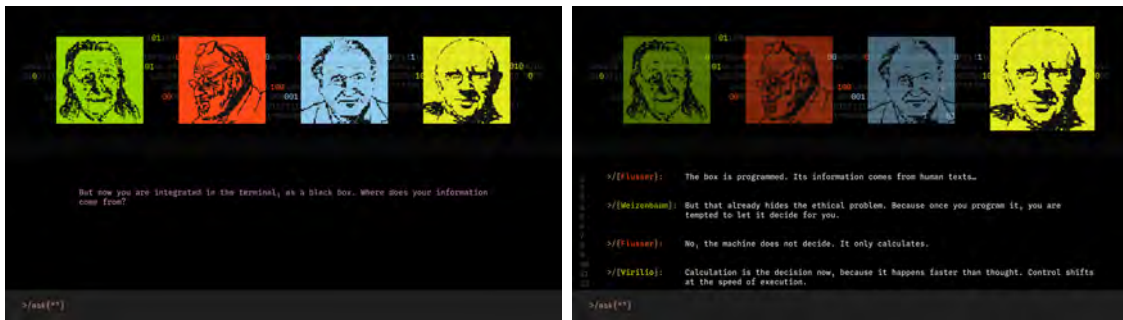


Figure 5.7: Concept 1: Portraits in a row with decorations, dither style portraits, non-static user question, minimal text output, bottom-to-top flowing text



Figure 5.8: Concept 2: single AI-philosopher portrait, ASCII style portraits, static user question, chunks of text, top-to-bottom flowing text



Figure 5.9: Concept 3: Single AI-philosopher portrait with decorations, ASCII style portraits, non-static user question, chunks of text, top-to-bottom flowing text

5.2.3 Group Interviews

The users in the first formative user tests were all interaction design student, with different backgrounds providing expert knowledge needed to evaluate the product, see table 5.1.

The data from the group interviews in the early formative user tests resulted in five overarching themes: *Text content*, *Dome Visualization*, *UI - visual elements*, *UI - text display* and *Transparency and Trust*, and three sub-themes: *Atmosphere*, *Flow of Interaction* and *User Interaction*, as well as data being out-of-scope: *Attitude toward AI usage*, as seen in figure 5.10. The user feedback gathered in these themes is summarized in this section. The quotes are translated from Swedish to English.

Table 5.1: Group Interview participants

Participant	Group	Educational background	Current field of study
P1	G1	Industrial design	Interaction design
P2	G1	Cognitive science	Interaction design
P3	G2	Industrial design	Interaction design
P4	G2	Software engineer	Interaction design



Figure 5.10: Thematic analysis of group interviews

Text Content & UI - Text Display:

It was clear that users had trouble following the debate that was being held. Both P2 and P3 mentioned in different interviews that the AI-philosophers used a lot of confusing words, and they felt like they were not on the same linguistic level as the AI-philosophers. Still, when asked about simplifying the discussion, multiple users expressed that they would not prefer this as it would not feel like a philosophical debate, thus losing its authenticity. P4 felt like slower speech and pauses would make it more understandable. However, other participants stated that the AI-philosophers spoke too long and slow and gave no room for audience questions. All users wanted to ask follow-up questions, but felt intimidated to do so. One participant mentioned that having the AI-philosophers continuously repeating themselves with keywords would break the immersion of a real conversation, and make it more feel like AI generated text.

Transparency and Trust:

When asked about whether or not you would trust the output of the AI-philosophers as "real", P3 & P4 stated that they would not trust the answers given to their question by the AI-philosophers. P4 clarifies by saying that:

"No, I would not have trusted that this is the meaning of life. But that is not the purpose either. The purpose is that I should get someone else's answer on what the meaning of life is, without strong thoughts of my own. The question is asked to start a discussion rather than get a concrete answer, and then I don't have a problem that it is generative AI. Because they will not claim to give the answer [truth] instead just that it is a philosophical standpoint"

P1 & P2 discussed that they generally do not trust AI outputs. They also expressed that the general trust in the system would increase if they knew where the sources of information the AI models used came from.

Participants also expressed that it should be clear to the user that the AI-philosopher speaking is not the actual person, but an AI-model. P4 said that "It feels OK to make representations, like in movies, but it depends. Conceptually it is OK, but the users have to be aware that it's not the real person they speak to." All participants raised ethical concerns about the encapsulation of a deceased person into an AI-model, and felt it to be dangerous if it was made believable enough to convince people that it reasons from the real person. P2 said, in regards to interaction with AIs of deceased persons, that "[...] It's scary that one can do this, scary if you connect it to the deceased person."

Flow of Interaction & User Interaction:

Regarding the interaction, all four participants expressed uncertainty in whether or not they would want to voice their opinions in front of a live audience. One possible explanation for this is that speaking in front of an audience may feel uncomfortable, but it could also be related to the subject of the discussion. P2 said that "They are scary men, feels intimidating to interrupt. They are men who talk about very big topics." Moreover, in G2, P4 said that "It would have felt more relaxed and playful if there was a possibility to ask your question anonymously". The same participant mentioned that philosophy can be a deeply personal topic, and expressed concerns in expressing those opinions publicly.

UI - Visual Elements:

In both group interviews participants expressed a necessity of keeping the user question on the screen to stay in the context of the question being discussed. There were conflicting thoughts regarding the AI-philosophers text on screen. In G2 the participants stated that having a lot of text on screen, and keeping the text history from the other AI-philosophers, made the prototype too text heavy, and wanted to reduce a lot of the text and remove the history. However, in G1 the participants discussed that they wanted the amount of text and the history to essentially stay in the loop, being able to go back and read if they missed anything.

Several users confirmed the hypothesis that having color-coded names and portraits was good for keeping track of who is speaking and remembering the different AI-personas easier, no user said that it was unnecessary or overwhelming. P3 said that "Connecting the answers to an image or face is powerful. [...] being able to connect the images to a color and a thing [the AI-philosopher's topic of interest]."

Atmosphere, UI - Visual Elements & Dome Visualization:

As previously mentioned, multiple users stated that the atmosphere was not relaxed, and that they felt intimidated in interacting with the system. Regarding the background visualizations, they expressed wanting slow moving visuals in the same style of that of the UI. In G1 they described that they wanted something "Powerful, transparent and not-jittery". Participant P4 in G2 mentioned that, if you can clearly distinguish that it is background visuals paired with a separate screen, then it did not fully utilize the full-dome space, and makes it feel like it "might as well" have been presented in a "normal" exhibition space.

5.3 Prototype Implementation & Development

Following the thematic analysis, the implementation of the UI and visual elements into high-fidelity, functioning prototypes went underway. This included *Art Asset Development*, *Front-end Development* and *Supporting Back-end Development* as well as server implementation.

5.3.1 Art Assets Development

Multiple portraits styles were created based on the team's shared vision, these accumulated into four different alternatives as seen in figure 5.11. These four abstract representations of the AI-philosophers were created for future testing with users. Three of them, set 1, 2 and 4, were created by us, and the third one was created by the artists MF&WS. The three sets were made through digitally hand sketching a black and white portrait base and online editing. The first portrait featured a digital ASCII-style rendering, with single colored typographical symbols on black background. The second portrait was of higher detail, using a dithering effect with black squares on colored background. The fourth portrait was the same render as the second portrait, swapping the foreground and background colors. Portrait set 1, 2 and 4 had corresponding GIFs which acted as talking animations.

Based on user feedback, a calming background of slow-moving stars made up of typographical symbols was created to match the UI. The domemaster format was 360-visuals rendered frame-by-frame to create a short moving looping sequence. Here, the AI-philosophers take the form of constellations in the sky, see figure 5.12 for development and mock-implementation through a domemaster simulator.

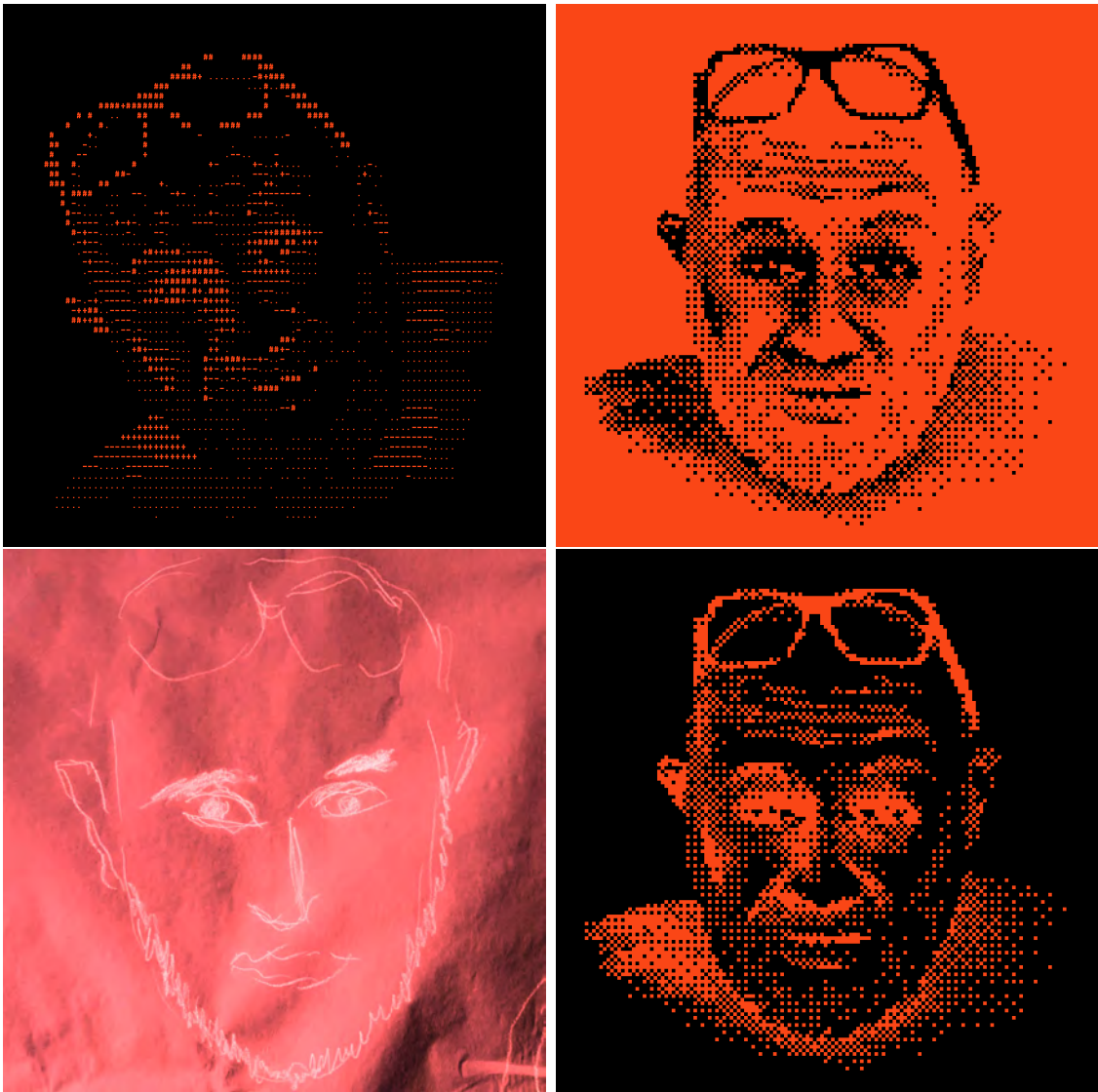


Figure 5.11: The four style alternatives of AI-Flusser's portrait

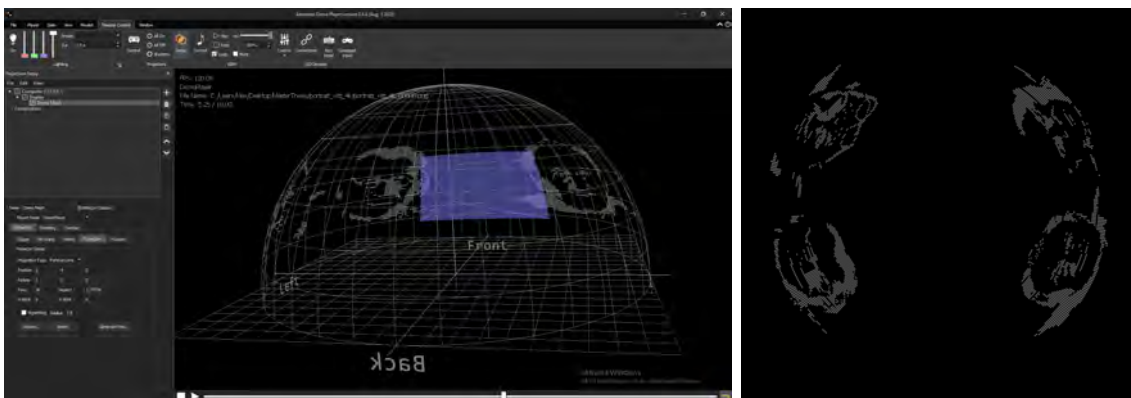


Figure 5.12: 360 full-dome format background visual and mock-implementation

5.3.2 Front-end Development

The front-end of this project was implemented as a modular React application using TypeScript, combined with CSS for styling. A mock server from the back-end side was built, so that the front-end could be tested and iterated using pre-generated responses from the LLMs. The UI was divided into four main architecture points; an image grid that represents the AI-philosophers, a voice interaction function for the user, a discussion log that shows the history of the debate and current user question, and a menu function to control the back-end system along with some UI elements.

Each AI-philosopher and its output are distinguished by color-coded names and corresponding images, with animated portraits for image set 1, 2 and 4 indicating when an AI-philosopher is speaking. The images is placed on an image grid, with all four philosopher portraits placed in the same row, see figure 5.13



Figure 5.13: The UI projected in Wisdome with overlapping background visuals

The front-end presents a central discussion area where AI-philosopher responses are displayed. The output text is placed at the bottom of the screen, moving upwards as more text is generated. Only four lines of text are visible at a time, and only the active AI-philosopher's text is visible on screen, meaning no history of the debate is visible at any time on the screen. The third and fourth (top) lines have lower opacity than the rest of the text, creating a transition as the text fades up out of view. The name of the AI-philosopher speaking is statically placed to the left of the first (bottom) line of text, switching names when the next AI-philosopher starts talking. The output text uses a character-by-character typewriter effect to imitate generative AI speech, making it appear as real-time generated text. The output text was fetched from the backend server through an API.

The transition between speakers is highlighted through a text-based thinking indicator that provides a text clue about the flow of the conversation. The interaction was synced with a delay that showed the upcoming AI-philosopher loading before the next response started generating. This was done through fetching the queue from the back-end, since the responses had finished generating in the backend far quicker than they were displayed in the front-end. This was to get a more fluid feeling of conversation between the AI-philosophers, and to lessen cognitive load for the user.

Furthermore, user questions can be sent in either by typing or holding the *spacebar* key to activate a push-to-talk voice input, with real time text-feedback indicating that the system is listening. Once the question is submitted, it is portrayed in the middle of the screen, between the AI-philosopher images and their output text. Speech-to-text was integrated using the browsers native speech recognition. The system also mutes the AI-philosophers conversation during voice input, to simulate the AI-philosophers listening.

Lastly, the concept incorporated a menu function meant for the team to control the back-end server and front-end display through functions such as: *pause*, *stop*, *fast-forward*, *swap image set*, *toggle intro sequence* and *display credits*. The different functions, as well as toggling the menu were are mapped via hotkeys on the keyboard.

5.3.3 Supporting Back-end Development

The back-end was built in Python and provides the main logic for the AI-philosopher debates. This was created by student Tengman under supervision of Scionova. It handles all communication with the LLMs, manages the flow of conversation, and organizes the contextual information each AI-philosopher uses to respond. The back-end utilizes a REST (representational state transfer application programming interface) API that the front-end uses to send questions and receive answers for the AI-philosophers. When a question is submitted, the back-end communicates with another LLM, prompting relevant background material for each AI-philosopher. This background material is selected from a curated set of documents and texts, which means personal descriptions about the provided from MF & WS, as well as additional material containing information from online sources. These documents serve as the systems training data, in order to get relevant responses. The backend then returns each AI-philosophers answers to the front-end, simulating a debate, see figure 5.14 for a visual description of the current human-machine system.

Creating a dramaturgy in conversation and dynamics between the AI-philosophers was based on the results of the co-creation sessions and user testing, as well as personal anecdotes from MF & WS, who knew the philosophers when they were alive. These co-creation sessions laid the foundation for design decisions regarding fluidity of conversation and character believability.

Firstly, due to users not fully comprehending the dialog, while not wanting to simplify the language, a commentator/moderator role was appointed to the "system", which appeared on screen summarizing the core points of the discussions after each debate. This was thought to make it easier for the user to follow the debate. The

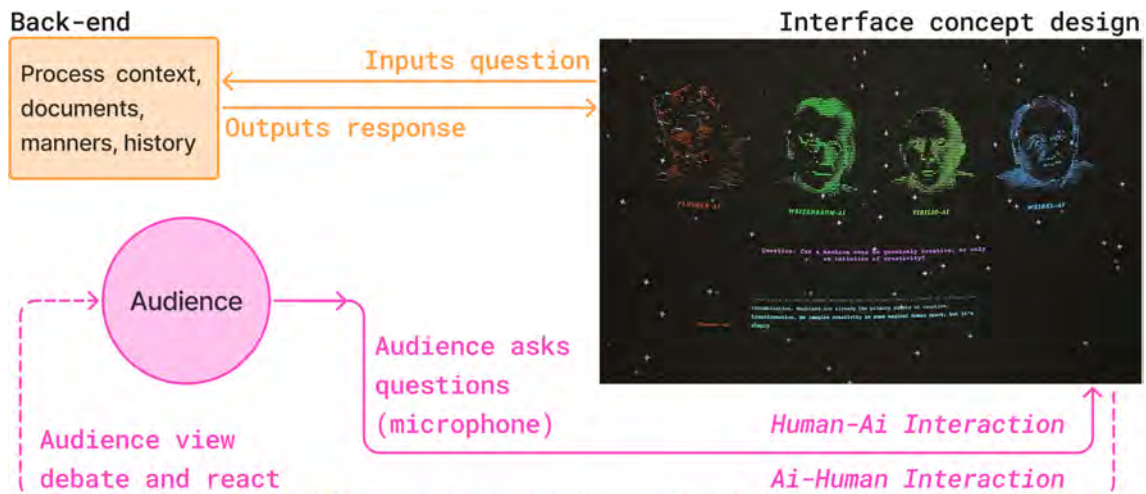


Figure 5.14: Simplified human-machine system, where pink describes the audience interaction, orange the back-end system process and the UI to the top right

AI-philosophers were also tweaked to use more every-day sentence structure and alternate their wording when appropriate.

Secondly, with the product now further along, the team established new milestones. These goals were to try and achieve more believable characters, and to create a more fluid and engaging conversation. Both were brought up during user testing as important for an engaging experience. While users had brought up ethical concerns about recreation of believable characters, the front-end clearly communicated that these were AI through renaming them "*Surname-AI*" and through the computer aesthetics.

Lastly, the team decided on testing two different dialogue modes for the system. The two were a debate mode and a panel discussion mode. The debate mode was thought to be more heated, with disagreements and disputes, potentially being more vague in answering the user question. The panel discussion were more direct answers to the user question, one by one. These two different dialogue modes were implemented for further user testing to discern what would be the most engaging, and most inviting for asking questions for the user.

5.4 Audience Evaluations

Having a complete functioning high-fidelity prototype, it was time to test it with live audience in the exhibition space. These tests are called Wisdome User Test 1 & 2 and are described in the following sections. Both were formative, and informed future design decisions and the design iterations.

5.4.1 Wisdome User Test 1

The test was structured in order to evaluate the four different portrait styles, one background-visual style, two different server settings (debate and panel logic), the

overall LLM dialogue and characteristics, audience interaction and curator role. See figure 5.15 showing the interior of Wisdome during the first user test. The event hosted around 40 audience members and collected 24 survey responses which will be presented in section 5.4.2.



Figure 5.15: Images from the first user test in Wisdome

The test was initiated by tech preparations before audience arrival, and the event was then initiated with an introduction to the project, whereas the AI-philosophers introduced themselves. The four different portrait sets were toggled in between different event states. Portrait set 1 for the debate logic and predetermined question, set 2 for audience questions, set 3 for panel logic and predetermined question and finally, due to the lack of time, the final portrait set 4 was previewed after the event was finished. Portrait set 1, 2, 4 had corresponding GIF-sets signifying when the AI-philosopher was speaking, portrait set 3 did not.

The first predetermined question was: "Can a machine ever be genuinely creative, or only an imitation of creativity?" Which the first AI-philosopher, AI-Flusser, gave an extensive answer to. Following this, all AI-philosophers answered with about two sentences each before the voiceless "System" prompted them further based on what they had discussed, and the debate continued, now more dynamic as AI-Virilio came back to answer multiple times. This debate sequence took 8 minutes and was manually interrupted by the control room as they did not stop generating answers.

Following this, the server was restarted and an audience member was then invited to ask a question. They asked "How do you define intelligence?" The speech-to-text API did not register what they said, and prompted: "How do you define a sentence?" Asking the question again, it read: "How do you define anything?" This was followed by the control room instead writing the question manually. AI-Virilio answered the question by saying:

AI-Virilio: "Intelligence is not improvisation. It is disappearance. The technological system does not negotiate with human consciousness. It eliminates it at computational speed."

AI-Weizenbaum answered next, referencing AI-Flusser despite him not yet having spoken about this:

AI-Weizenbaum: "Intelligence is not a performance or a disappearance, but a profound human capacity for moral judgment that cannot be re-

duced to computational processes. When Flusser and Virillio both treat intelligence as something [...]"

This debate was also manually interrupted after 8 minutes, followed by the back-end system being switched to panel-mode. Due to lack of time, only one pre-determined question was asked, after 7 minutes of discussing the last question the panel was manually interrupted in order to have time to fill out the survey.

5.4.2 Survey Results from Wisdome User Test 1

The survey got 24 responses and the demographic was represented by AI usage. A majority of 58% used generative AI tools daily, while 38% used generative AI a few times a week and only 4% a few times a month. There was a generally even distribution on the cause of AI-usage; a multiple choice questions that resulted in *Information seeking & learning* (38%) being the most common usage. Following this: *Writing & communication* (24%), *Problem solving & technical assistance* (18%), *Creative content generation* (15%), *Personal productivity & planning* (10%), *Other* (1%). There was an even distribution of positive and negative, regarding the question about the audience general perception of generative AI, see figure 5.16.

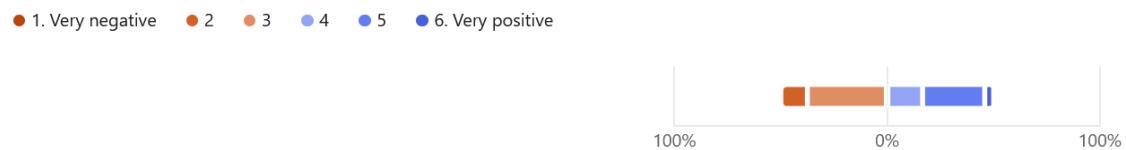


Figure 5.16: General position toward generative AI tools

Portrait set 1 and 4 were preferred by the participants, both receiving 33% of the votes each. The panel logic received 75% of the votes, however the panel logic was denoted as the 1st logic, which in reality was presented as the debate logic, which received 25% of the votes, making the results inconclusive.

The participants were asked to rate their experience in three words, in a free-text field. The most common one written was *interesting*, notably, people also described the show as: *slow, confusing, fun, boring, repetitive, new, surprising, reflecting/thought-provoking*. When rating their experience of *content believability, conversational fluidity, and visual elements*, it was clear that the conversation was not perceived as fluent, instead rather machine like. The believability and visual elements were rated comparably fairly high. There were however, for both instances, more than 30% of participants who put a score ≤ 3 . See figure 5.17 for Likert scale results.

As the AI-philosophers discuss different perspectives, such as AI-ethics and moral accountability, institutional power and a systemic view of human-to-technology usage, and the speed of technical development the survey finished with a tangent question of the themes discussed. When asked if the participants felt responsible for the outcome when delegating a choice to a machine 87% said yes, and 13% said no, see figure 5.18

5. Results

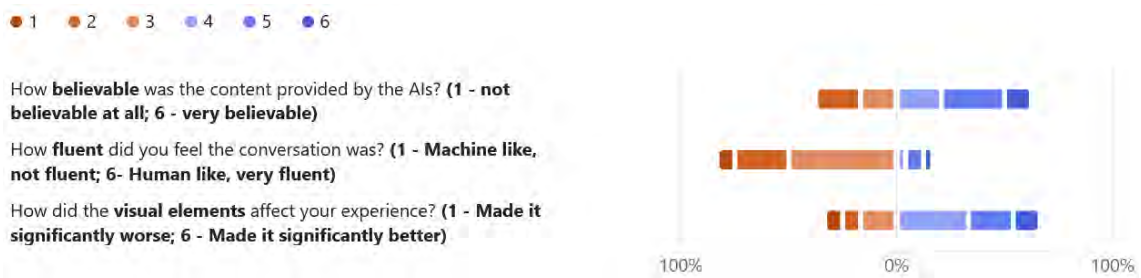


Figure 5.17: Results of rating the experience based on statements

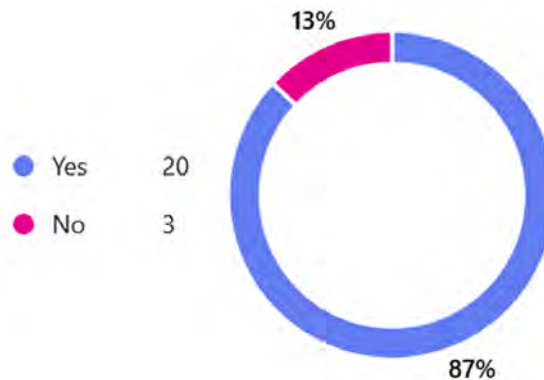


Figure 5.18: Survey answers regarding accountability

5.4.3 Thematic Analysis of Wisdome User Test 1

A thematic analysis revolving the first Wisdome test was performed. It contained data from the survey responses, an interview with an expert user researching creative AI development and stakeholder meetings and team discussions. The data got transferred into post-it notes, and was organized in different themes, which were: *Personalities and the role of the Moderator*, *Content and Dramaturgy*, *Tempo*, *Visuals*, *Sound and Ambiance* and *Audience Interaction*, see figure 5.20 and 5.19 to see the overall structure and process of the thematic analysis. The themes are described in this section.

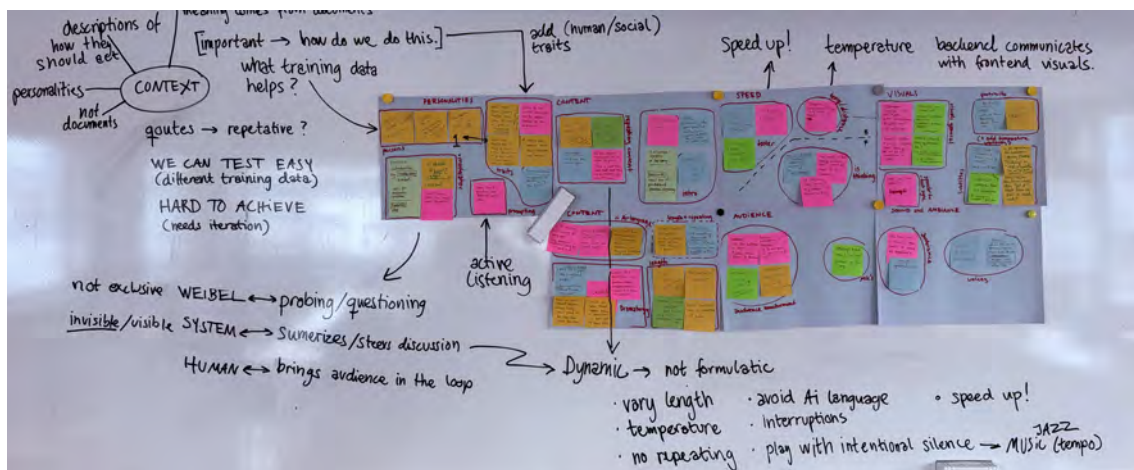


Figure 5.19: Thematic Analysis of back-end themes

Personalities and the Role of the Moderator

Many participants mentioned that the personalities were perceived as rigid and lacked human nuance. This was thought to be because the AI-philosophers often referred to their main talking points, e.g. Virilio-AI re-framing every answer to concern "speed" and "acceleration", or the AI-philosophers re-using sentence structures. One participant wrote: "rigid personalities more interested in conflicting each other than going somewhere with the conversation." There were also discussions surrounding the AI-philosophers lacking human traits such as humor, sarcasm, emotions, and not being anchored in real world examples. Some participants expressed confusion as they were unsure who the AI-philosophers were, even though they had introduced themselves in the beginning of the show.

As expressed by the audience and as a reflection from the team, the moderator role feeling unclear. One participant wrote "Maybe add an AI conferencier? Or is it supposed to have a human conferencier for each viewing?", another wrote that it was "Confusing and difficult since the replies were difficult to follow and it was not very easy to follow the replies or have it summed up." During the test, the "system" summarized the discussion in text. However, the "system" had no voice and did not get an introduction like the AI-philosophers. More clearly defined roles were needed for next iterations.

A participant brought up the topic of ethics in the survey, questioning how ethical it is to re-create deceased people through AI, they wrote:

"Additionally it raised the question in me on how ethical it is to in a way impersonate a person in this way, since an AI is interpreting their recorded thoughts and creating new takes based on it instead of their actual thoughts."

Another participant brought up a different perceptive, of how this use-case of generative AI can mitigate current risks:

"It highlights ways to use ai to get information from different perspectives. One of the biggest risks with using ai is lack of nuances. This way of using ai can mitigate that risk with different viewpoints being presented in a balanced way."

Content and Dramaturgy

In the free-text fields six participants reported the conversation as difficult to follow due to complex language, one participant wrote that "It was challenging to understand their arguments when they use very advanced English language." There was also feedback saying the immersion was broken when the AI-philosophers started repeating themselves, making it feel less like a debate. Their sentence structure was also perceived as artificial, as one respondent wrote "The agents easily got repetitive, kind of like no its not what you say, its what I say all the time. [...] They also used a lot of typical AI language structures." Another participant also pointed out this language structure: "The philosophers seemed more interested in contradicting the others than seeking some kind of deeper knowledge."

5. Results

The overall flow of conversation was thoroughly discussed within the team. The lack of synchronization between subtitles and speech was perceived as distracting and took away from the immersion. The answers were also lengthy. One participant wrote that: "The flow of the convo is the weakest part, the agents dont hook into each others talking points, they just pile them on top of each other." Overall, there were many participants who stated that the dialogue sparked curiosity to learn more, but was described as an interesting concept that is in need of work.

Tempo

Many participants stated that the tempo of the dialogue was too slow. There were a lot of pauses, and the AI-philosophers spoke too slowly. One user stated that: "The concept is very interesting, however the execution need to be quicker to keep attention."

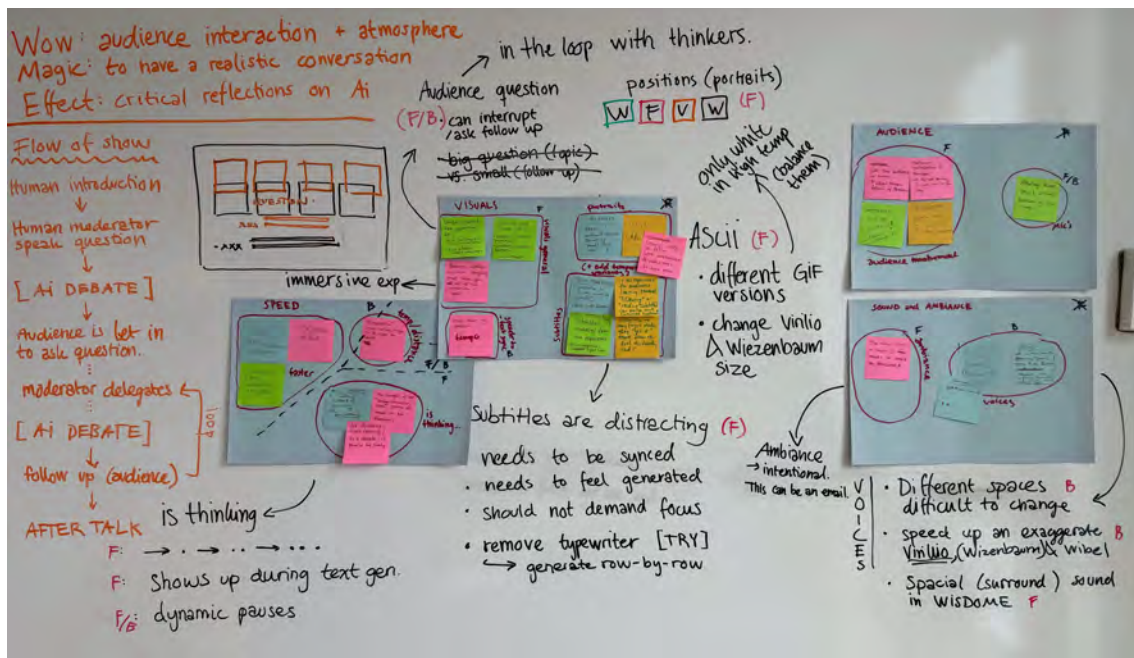


Figure 5.20: Thematic Analysis of front-end themes

Visuals, Sound and Ambiance

Regarding the visuals in general, including both the portraits and the dome visuals, the audience had rated the visual element as something that heightened their experience, see 5.17. One participant commented that music would heighten the experience. Four participants specifically mentioned the aesthetics as a positive, one participant commented on the retro-aesthetics as creating immersion "I think it was a good combination between real and computer, it made sense." However, several people also noted that they wished the space would be utilized more, one participant wrote that:

"Another thing that I think is worth noting is the size, I feel like having them so low down in the globe and quite narrow is a missed opportunity. Obviously, if it does not work, then keep it this size, but otherwise maybe

work to make it larger and more immersive. E.g. almost having to turn your head to see the different people"

Regarding the portraits, there were three comments pointing out that the talking animations for portrait set 2 and 4 seemed artificial "The visual version where the mouths moved felt quite artificial." and "In regards to the visual elements, I feel that version 4 is the best visually, but the animation is so creepy in nr 4 (and for nr. 2 too) so it ended up that I still prefer nr 1." The third user also specified that the lack of movement such as for portrait set 3 felt *useless*: "Lastly the visual element for nr2 (of the 4) gave uncanny valley and was more distracting and not inviting. The third had no moving elements so felt rather useless." Another participant said that it was hard to map who was speaking when there was no movement: "A voice without a visual is harder to follow and less plausible. Interesting for me to realize the different visual influenced my willingness so really take into account what they said."

While the movement in portrait set 1 was preferred by users, the placement of AI-Weizenbaum (green) and AI-Virilio (yellow) next to each other made it hard to differentiate due to similar colors. One user pointed out that the colours were difficult to differentiate "the green and the yellow are too similar to each other."

Audience Interaction

The system was generally not perceived as interactive. The audience did not have many opportunities to interact with the system due to technical errors and time constraints. There were also no instances of interruptions or follow up questions, and all questions were directed to the entire panel of AI-philosophers. One participant said: "I would like to talk with them individually, maybe to have a follow up question."

5.4.4 System Design Iterations

Based on the results from the thematic analysis, several key design decisions were made through co-creation sessions. These regarded both back-end and front-end development as well as aesthetic improvements.

Back-end Conversation Dynamics

To counteract the perceived repetitive and uninteresting dialog, there was a need to update the LLMs contextual backgrounds with personality traits, descriptions on how they should act as their persona, as well as human/social traits such as humor and sarcasm. The AI-philosophers were also prompted to bring up real world examples, with the goal to anchor them and let audience relate to them more. They were also tweaked to use less artificial and AI "typical" language structure. This meant using varied length of responses, avoid repeating common words, phrases and sentence structures. This was thought to be achieved through keeping the history of the conversation in the back-end.

To achieve a more fluid conversation, a relevance tracking system was implemented in the back-end. This was done through a "distance" metric, where a low distance

meant that the question or previous response was highly relevant to the current AI-philosopher, and a high distance meant that it was of low relevance to the AI-philosophers. Through this distance metric, several design decisions were made regarding conversation fluidity, relevant to the back-end: The metric decided who would answer first, and the AI-philosophers could now interrupt each other if the distance was low enough. This led to the interrupted AI-philosopher then answering back. The interrupt mechanic was also intended as a way to create unpredictable, novel events. Moreover, The different voices and speech speeds were also altered based on the audience feedback of wanting them faster and more balanced.

The moderator role was re-imagined in three parameters. One of the AI-philosophers, AI-Weibel, was given the moderator role, and was promoted to briefly summarize after a discussion was finished, as well as to probe and challenge the views of the other AI-philosophers. AI-Weibel was chosen as the moderator due to him not being present in the original artwork, Home of the brain. Thus his character was thought to have an outside perspective compared to the other philosophers, as well as his core beliefs about institutions and power being fitting for a moderator role.

The previous moderator/commentator, "system", was given a role invisible to the audience, with the purpose of delegating the word and organizing the back-end queue. The final role was given to the human curator, who was to bring audience in-the-loop and focus on managing interactivity.

Front-end Display Changes

Several issues were addressed, such as; how to sync the tempo of LLM's voice-lines with the on-screen text generation, improving the contextual recognition of the push-to-talk function recording user questions, adding a menu to the interface for operator ease-of-use, and so forth.

Through a token system implemented in the back-end, the front-end could now sync the text generation with the length of the generated mp3 voice files that were fetched from the back-end. The subtitle flow was also updated as it was seen as distracting by users. The typewriter effect was removed, and instead the subtitles now mimicked the mental model of subtitles for movies, with two static lines. With a line-break after a full-stop, or equivalent if not present. The new subtitles functioned as an accessibility aid while still allowing the audience to focus on the voices and visual experience.

The distance metric was communicated to the front-end through the back-end and provided more dimensions to create dynamics and novel events. A second and third GIF set was introduced for the chosen portrait set, set 1. The first new GIF set was a "thinking" animation, where the portraits faded in and out slowly. The second new GIF set was an *highly distorted* "talking" animation. The normal "talking" GIFs triggered at a high distance threshold value, however, for a very low threshold value the highly distorted talking GIF was triggered instead. The timing of the thinking-delay was now also dynamic, where a low distance had shorter "thinking time" than high distance responses.

New State-changing Functions

Analyzing the user interaction and event-flow, several state-changing functions were defined to create a seamless interplay between the front-end and back-end. These were defined through their relevance in different use-states in an updated Story Mapping: *0. audience enter; 1. ask new question; 2. debate; 3. follow-up; 4. end.*, where the system loops between *3. → 1.*, or *3. → 2.*. Each function was prioritized using the MoSCoW framework and can be read in detail in appendix A.9.

Through this analysis it was decided to implement several new state changing functions. These were to be accessed through a separate front-end controller interface that would be used from the control room as well as through a small portable keyboard in the hands of the event curator. This control interface included the functions:

- **Round-up** - Forces AI-Weibel to be the next AI-philosopher in queue, rounding up the debate and setting idle state "new question".
- **Pause** - Immediately pauses when the current AI-philosopher finishes speaking and setting idle state "new question".
- **Deactivate Gifs** - Manually reverts the image grid to the portrait sets if the system bug occurs
- **Hard Reset** - Clears current question, AI-philosopher queue and sets idle state "new question" without clearing server history.
- **Clear screen** - Visually clears current question
- **Un-mute / mute** - When un-muted the system picks-up any voice input as a new question, has to be manually toggled by curator.
- **Boot-up sequence** - sets starting screen visually "booting" the system.
- **Credit screen** - Sets Credit screen showing credited names and a survey-QR code.

5.4.5 Wisdome User Test 2

A second, shorter, on-site test at Wisdome was held with two audience members. This was orchestrated by the the front-end and back-end team in order to evaluate the new changes made as well as test the new flow of interaction with regards to the new state changing functions. The two audience members represented different demographics in the broad target group, see table 5.2.

Table 5.2: Wisdome User Test 2 participants

Participant	Education	Current occupation	English lvl
P5	Software engineering	Industrial design student	High
P6	Non	Early retiree	Medium

The test included initial questions from the curator as well as an instance of interrupting the AI-philosophers with a follow-up question, in order to on-board the audience. The test also included audience questions, the curator manually muting and un-muting the system through the portable controller and giving the audience the microphone. One audience question included an attempt to ask a singular AI-philosopher a question, see figure 5.21 to see images from the test.



Figure 5.21: Images from the second user test in Wisdome

After initial debate of the opening question the curator attempted to interrupt which led to a bug in the system. Trying to see if the server had crashed, a random question "what do you think an apparatus is and how does it define us as humans?" was asked, leading to some humorous interactions after some time debating, as AI-Flusser exclaimed in a angry voice:

AI-Flusser: "Stop right there Weibel! You assume that data allows for protest, but the apparatus has already programmed the very language you use to complain. If our humanity is just a series of pre-set options like the choices of a microwave, is there any difference between a judge and a kitchen appliance?"

AI-Weibel: "The apparatus is now a kitchen appliance. We have reached peak media theory. Flusser acts like being a kitchen appliance is a bad thing, but at least a microwave delivers a predictable result for the public while your human judge is a black box of secret, elitist whims".

This debate went on for 6 minutes. As they finished speaking, the UI showed "[Waiting for new question]" signifying that the system was ready for new questions. The audience members each asked one question, starting with the first participant asking "What would we do when people start claiming that real stuff that happened is AI generated?" speech-to-text API misheard the question two times, and it was correctly transcribed the third time. The second participant wanted to ask a question but felt their English was not good enough, so they communicated their question to the curator who then asked "Weizenbaum, when you created ELIZA, what was your thought process behind that experience and how did you react when people started getting emotional with the AI you had created".

5.4.6 Survey Results from Wisdome User Test 2

While the many improvements greatly heightened the experience, the test still identified several issues and bugs with the system that the back-end team needed to address. The overall experience was examined in a short survey of 9 open-ended, free-text field, questions.

Describing their Experience

The first participant (P5) had a rather negative view on AI, and used generative AI a few times a month. They described the experience as "Cool, interesting, impressive". Clarifying that it felt "cool" because of the aesthetics, "interesting" as it was a unique experience, writing "Interesting because listening to 4 AI-philosophers debate about AI is not your ordinary Monday morning activity! And is definitely not the common AI interaction that we usually have." And finally, "Impressive because the conversation between the philosophers was so well done! How they get angry at each other and answer each other."

The second participant (P6) was very positive toward generative AI tools, they answered all questions in Swedish, so all quotes are translated from Swedish to English for the purpose of this thesis. They described the event as "engaging, informative and entertaining". Choosing the word "engaging" as the interaction captured their interest making them want to continue the conversation. "Informative" as they felt as they learned something new and reached a deeper understanding of the topics discussed as the conversation progressed. Finally they motivate "entertaining" as it was not only a learning experience but fun and easy to get "hooked".

Believability

Regarding believability, P5 felt the interaction was believable as the AI-philosophers did not say anything unreasonable. P6 meant that the different personalities and expertise shone through, making them read as believable. And continue on by writing that the AIs interaction with each other heightened the experience. Exemplifying how the AIs cut each other off and gave new perspectives to each topic discussed. As P6 was present for the first test, see section 5.4.1, they compared the experience saying that the conversation now felt more spontaneous and was perceived to be less programmed, as the AI-philosophers now only interjected when they had something relevant to add.

Fluidity

P5 felt this dynamic added to the fluidity of the conversation, saying that it felt more human like than machine like as the AIs were responding to each other and commenting on things the other AI-philosophers had previously said. Additionally, P6 meant that the conversation felt fluent and dynamic as it was faster, more frank and sharp than during the first test. They did note that there was still a hint of the "programmed structure" behind the interaction, but their overall impression was natural and engaging. They also state that the added humor made it feel alive.

The change of being able to address a single AI-philosopher or the entire panel was appreciated by P6, they also appreciated that the AI-philosophers now kept building on each others statements, continuously growing the conversation in new directions. They describe that the debate sometimes created a "passionate polemic" which contributed to the dramaturgy of the event. P6 continue by comparing to the first test, commenting that the language was easier to take in and understand. While being generally positive toward the show, P6 wished to have the AI-philosophers introduce themselves instead by the curator.

Visual Experience

Regarding the visuals, both P5 and P6 felt that the visuals heightened the experience. P5 said that it helped them identify who was currently speaking. P6 said that the portraits together with the voices made the AI-philosophers be perceived as individuals. They continued on by saying that the AI-philosophers personalities clearly showed, and that they each fit with their visual representation while clearly communicating that they were AI through the computer aesthetics, without being distracting. Both P5 and P6 commented on the subtitle style being easy to follow and contributed to the experience as a helpful aid.

Interaction with the system

As the test only included two participants, both got the opportunity to interact with the system. P5 described this interaction as "unreal but also very cool, very enjoyable experience". And continued on by saying that it was helpful that the curator started by asking the system questions, as this gave an example of what type of questions you can ask. P6 described the interaction by saying "It felt like taking part of a discussion rather than using a common [generative AI] tool".

5.4.7 Final Design Iterations

While the participants confirmed that the system was on the right track, there were still some back-end issues and aesthetic changes needed to be made. In the final week of production, the back-end engineer implemented a new LLM to power the AI-philosophers. While this lost some of the humor and social traits, it allowed for better contextual description and more nuanced inclusion of real-world examples. Portraying different emotions based on the distance metrics was also added to the voice and text.

Regarding the text-to-speech recognition system, none of the participants mentioned it as an issue, being aware that it was still in a prototype stage. However, the team felt that it still left much to wish for, to make the interaction more seamless. The first participant repeated their question 3 times before the curator had to take over and answer for them, as the system misheard some words. This issue was addressed through implementing a separate LLM that analyzed the contextual clues of what was likely being said in the back-end, minimizing the risk of outputting incorrect words. This change prolonged the waiting time between the question asked and the AI-philosophers answering, but produced more correct interpretations of questions asked.

Final changes were made for the purpose of aesthetics and atmosphere. The looping video was made longer to not perceive any looping elements. The "shooting stars" and other moving visual elements were reconfigured to be more dynamic and less linear. Finally, the live-feed screen was mapped onto the background dome visuals, and in this space less background elements and "stars" now overlapped. Musician Jakob Dahl was commissioned to create a background soundtrack to be played during the final event.

5.5 Final System Design & Functions

This section describes the final product and the event in which the product was showcased, along with the results of the survey from the event. The final product functioned according to the visual description of the human-machine system as seen in figure 5.22. Where the front-end interfaces consisted of the UI visible for the audience and a hidden control panel; the environment controls influenced music, light and background visuals; the human-to-ai interaction was facilitated by a curator; and finally, the back-end communicated with the front end through a API web-server.

The figure describes four main systemic fields, the *Front-end interfaces*, *Environment Controls*, *Human-to-AI Interaction* and *Back-end* and how these interact and affect each other. In the following section these different fields, apart from the back-end will be described.

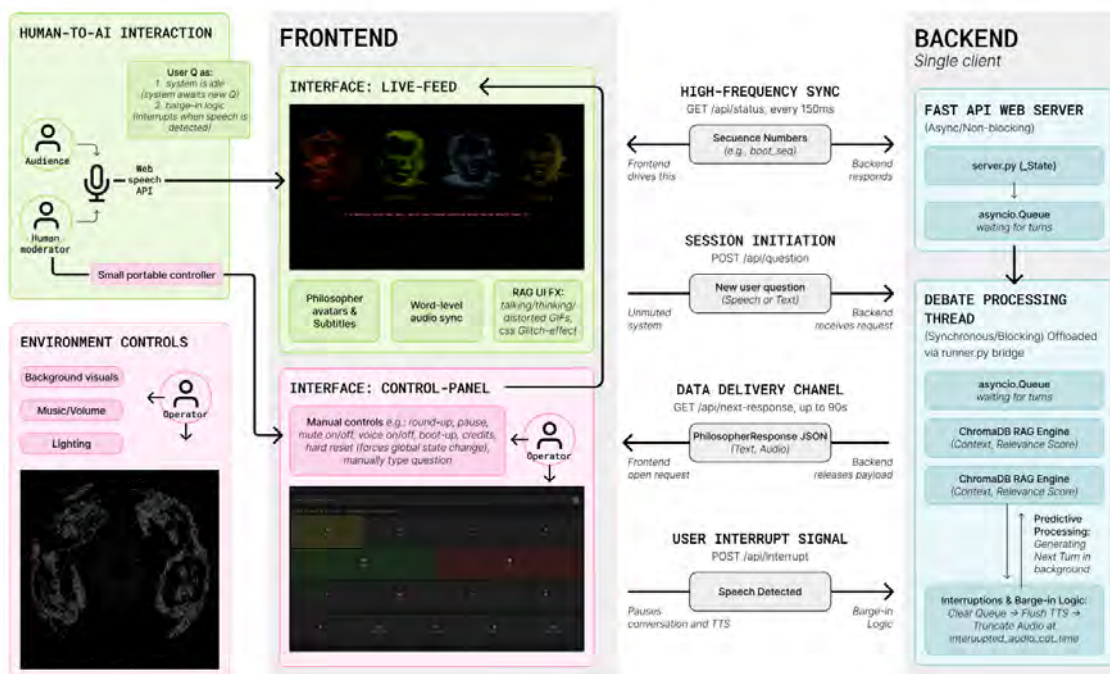


Figure 5.22: The final visual description of the human-machine system, the pink color represents controls not visible to the audience, the green color represents controls visible to the audience, and the blue color represent the back-end.

5.5.1 Front-end Interfaces

The front end interfaces were made separate to facilitate more control for the team during the showcase. The interface that the audience sees, *Interface 1: live-feed*, includes the portraits, user question and the subtitle system. *Interface 2: control-panel* is only available to the people in the control room and includes all state-changing functions described in section 5.4.4. These functions were also available through shortcut-keys using a small portable controller that the curator would be carrying.

Interface 1: live-feed

The final user interface features a retro-computer aesthetic, with a black background and primarily white colored mono-space text. The user question is hot-pink and placed in the middle of the screen. The AI-philosopher output text is placed in the lower part of the screen, with the AI-philosopher portraits placed on the top, see figure 5.23. The four AI-philosopher portraits are placed on the same row, with their names written as "*Surname-AI*" underneath each respective portrait, as seen in 5.24. For added contrast to the black background a "glow" effect was added to each AI-philosopher portraits.

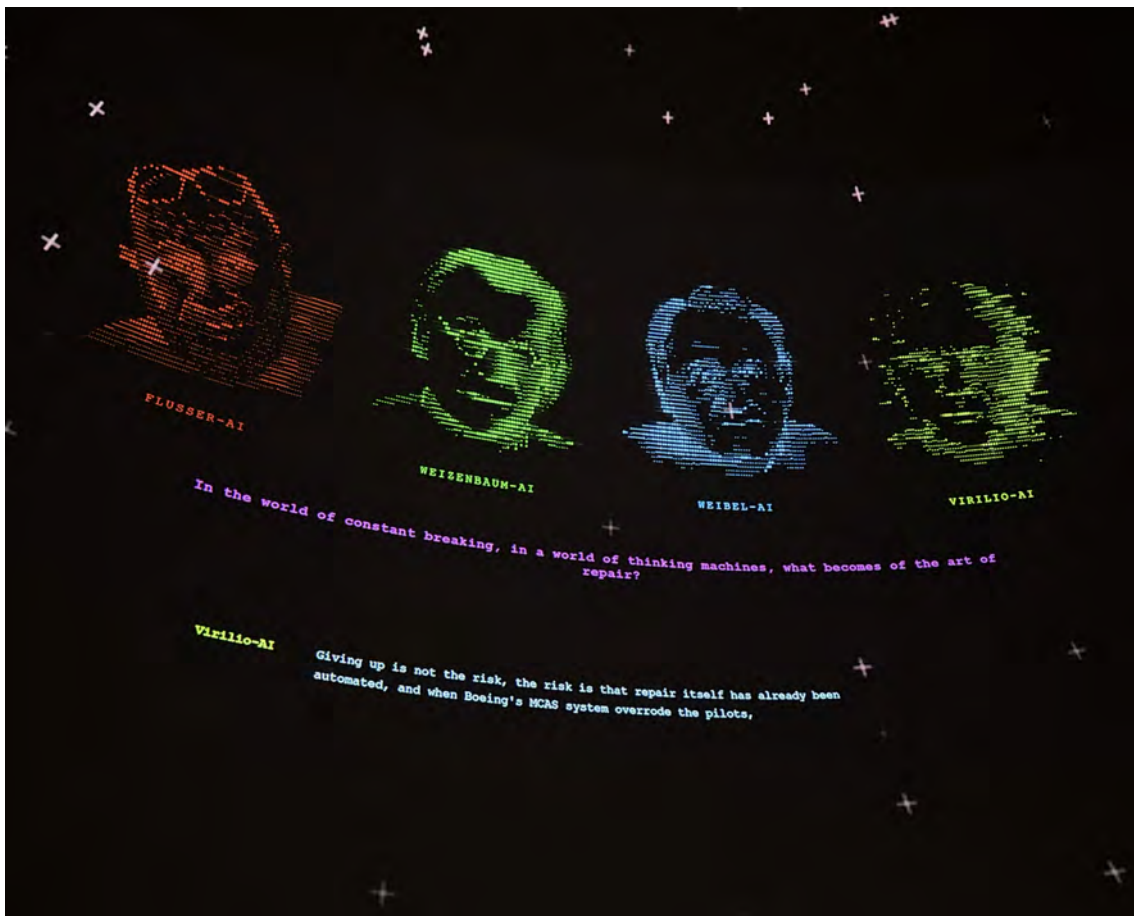


Figure 5.23: The final interface projected in Wisdome while AI-Virilio is speaking

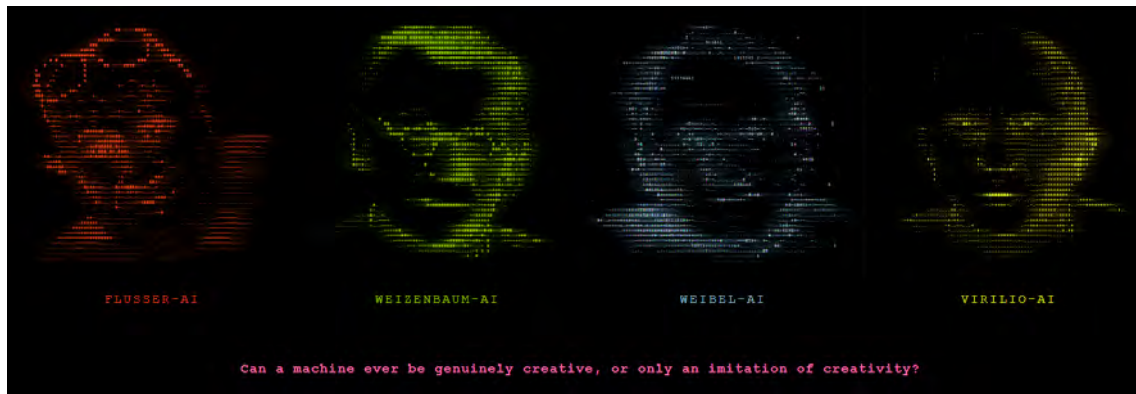


Figure 5.24: Portraits and user question while AI-Weibel is speaking, with low distance

When an AI-philosopher is speaking, two different GIF-sets can be played, one that is "calm" and one that is "extreme", based on the distance metric described in 5.4.4, that is derived by the relevance to the topic and the previous AI-philosophers thoughts. See figure 5.25 for some key frames of AI-Flussers normal, "calm" talking animation, and figure 5.26 for key frames from AI-Flussers low distance "extreme" talking animation.

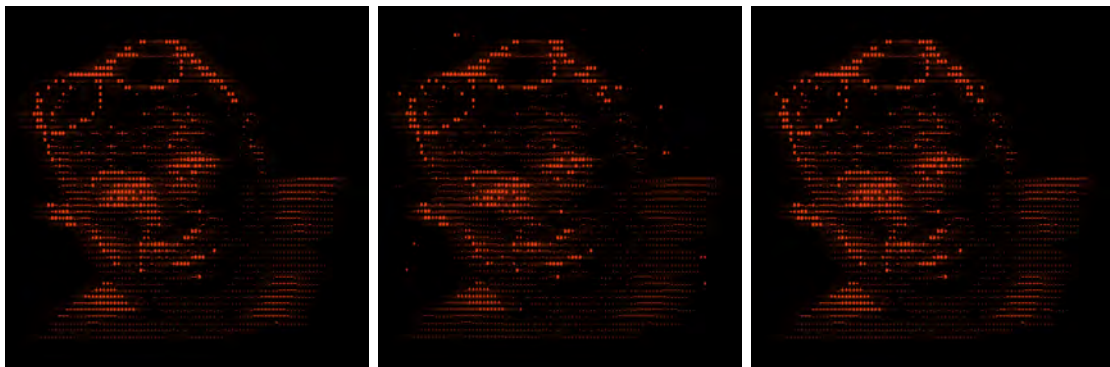


Figure 5.25: Key frames of AI-Flussers normal, "calm" talking animation

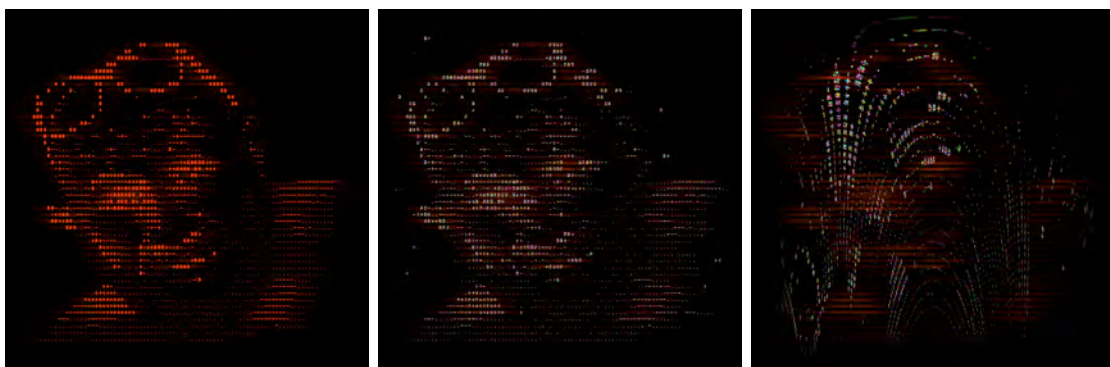


Figure 5.26: Key frames from AI-Flussers low distance "extreme" talking animation

Every AI-philosopher also had a "thinking" animation. The timing of the thinking animation is also based on the distance metric described in section 5.4.4. The thinking animation can be described as the AI-philosopher portraits slowly fading in and out of opacity in three steps, see figure 5.27

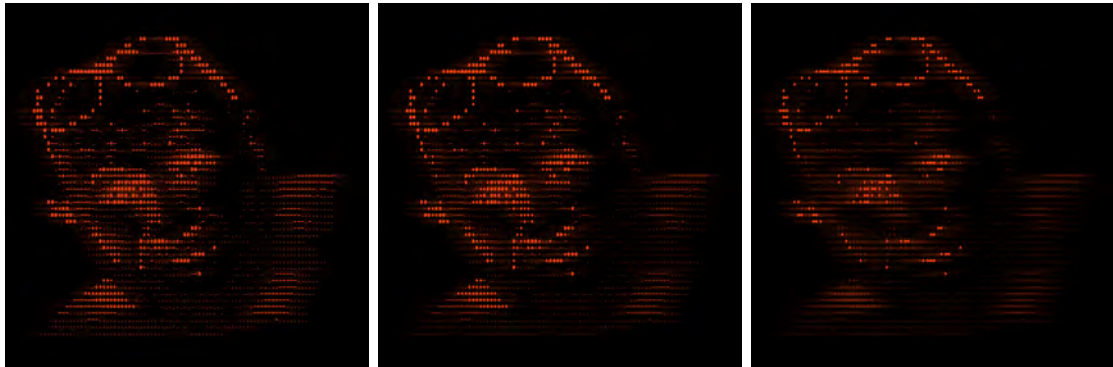


Figure 5.27: All frames in AI-Flussers thinking animation

Interface 2: Control-panel

The control panel, which functions can be seen in table 5.4.4 in section 5.4.4, is set-up on a different computer, and had manual control of both the user interface as well as some back-end functions, see figure 5.28. These controls are: *round-up*, *pause*, *mute on/off*, *voice on/off*, *boot-up*, *credits*, *hard reset* and the *ability to manually type a question*.

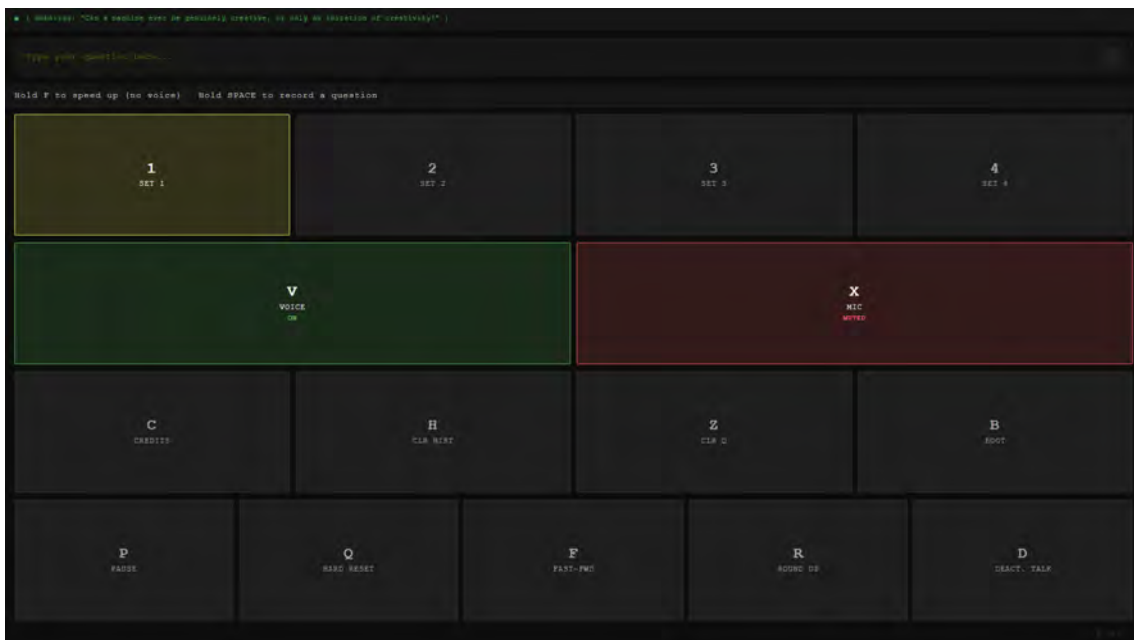


Figure 5.28: The interface of the control panel

These state-changing function such as the *boot-up sequence* can be manually activated through either via a hotkey on the keyboard or by clicking on the control-panel interface. The boot-up sequence is only a visual display informing users that the

system is powering up, in reality, the system is already online, See figure 5.29. In the same way the credit sequence can be accessed from the control panel, which showcases all the stakeholders and team members and contributors of the project, as well as a QR-code which takes the user to a survey, see figure 5.29.

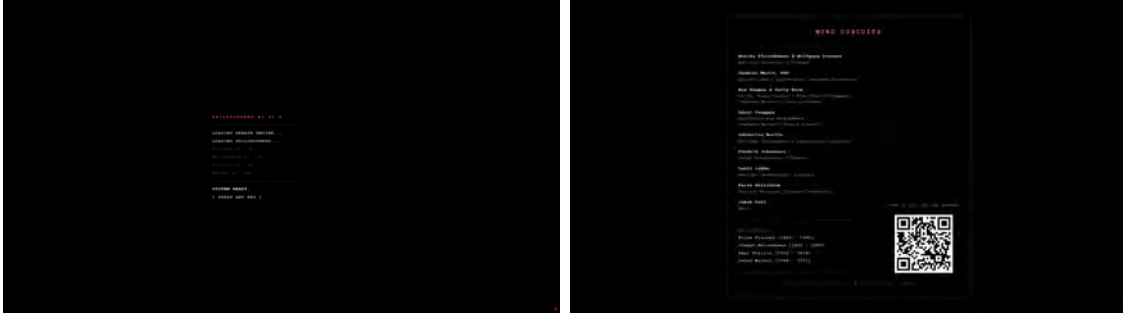


Figure 5.29: Boot-up and credit screens

The AI-philosophers output text now resembles that of subtitles, as described in section 5.4.4. The name of the AI-philosopher speaking is statically placed to the left of the output text, in corresponding color to that of the AI-philosopher.

5.5.2 Environment Controls

The environment controls allows the operator in the control room to manually set background visuals, play the music track and adjust the lighting based on the different event stages. In the final product the background visuals now utilized more of the space, with the portrait constellations taking up more space in the dome, and more effects being added. Those effects were shooting stars and randomized star shimmers. The amount of stars that overlapped with the user interface was also narrowed down to not take as much attention away from the content, but still make it connected to the interface for a seamless visual connection between the two, see figure 5.30.

During the show, an ambient musical loop was playing throughout the experience. The music is described as pulsating in and out, eerie, ambient, not hopeless while not necessarily hopeful.

5.5.3 Human-to-AI Interaction

The human-to-AI interaction uses a wireless microphone connected to the computer which projects the interface. The speech-to-text recognition system featuring a LLM that analyzed the contextual clues of what is likely being said in the back-end, minimizes the risk of wrong wordings. The curator has to manually mute and un-mute the system, through the small portable controller, to be able to speak to the audience in the mic without the system registering it as a question. In this way, the curator also has to un-mute the system when an audience asks a question. The un-mute function has a sensitive delay time while it was connecting to the LLM, during this time it is important to keep the microphone muted as the system crashes if it registers speech during this time.



Figure 5.30: Background visuals together with the credit scene in Wisdome

The questions can be directed toward a singular AI-philosopher by saying their name, or to the entire group if no direction is given. The questions can be asked when the system is in its idle state waiting for a new question, or while the AI-philosophers are talking triggering the barge-in logic.

5.6 Final Showcase

The production of the final showcase followed a pre-planned schedule designed to coordinate all involved parties and technical elements. The event hosted 57 people, and started with an opening statement and welcoming words from Universeum staff, followed by an introduction of the project and the AI-philosophers by the team.

Thereafter, the boot-up sequence was played and the show started with asking the system a predetermined question to act as an on-boarding sequence, showing the audience how to interact with the system via the microphone. The predetermined question was: "Can a machine ever be genuinely creative, or only an imitation of creativity?". Even though the system had been checked in advance, the AI-philosopher's voices did not activate. This required an immediate reboot. Following this, the curator asked "Does technology liberate us, or does it always end up controlling us? This question was answered by AI-Weizenbaum bringing up an example of algorithmic judges and who bears the responsibility behind the verdict enacted. The debate escalated with back and forth between all AI-philosophers, AI-Weibel whispered a summary to the audience, and the debate then continued. This question was debated for 3 minutes before the curator interrupted with a follow-up question, on-boarding the user on the interruption feature: "You talk about decisions we can leave to AI, but what choices do you think is ethical and moral to give to AI?"

A known bug appeared where, when asking a new question, the previous one would be generated on screen when the system was processing the new one. No other visible bugs were to appear during the entirety of the show. After the follow-up question the audience were invited to participate in the show by asking the AI-philosophers any question they would like. The curator would ask users to raise their hands if they had a question in mind, go up to them and instruct them on the microphone and when they were able to talk into the system. They would then ask a question, and the debate would unfold. If they did not succeed to ask their question, meaning if the system did not register the words correctly or they would be too slow for the system to register, the curator would repeat their question instead. See figure 5.31 to see the visual display during the showcase.



Figure 5.31: The AI-philosophers in Wisdome discussing the final question, AI-Virilio is currently speaking.

The first audience question was "How would we make the world the best place possible?" which led to a meta-commentary between AI-Weibel and AI-Flusser about the museum installation they were a part of and who chooses the questions they answer. After some further heated debate with the AI-philosopher interrupting each other the question was debated during 6 minutes.

The second audience question, which came from a child, was "What does the highest number in the entire world?" Which was answered in a humorous, yet serious way, by AI-Weibel, not knowing it was a child who asked it.

AI-Weibel: "The question is not innocent. "What does the highest number in the entire world?" is a broken sentence, and i will not fix it for you, because the break is the point. Who asked this question? Who approved it for this room, in this format, with these speakers? That decision already answers half of what we are here to debate. [...]"

This continued with AI-Flusser responding to AI-Weibel through pointing out that seeing the structural problem takes away the danger of it, AI-Virilio then responds with "Flusser says the invisible cage is more dangerous than the named one. Naming the cage is not the catastrophe i'm pointing at..." Before being promptly interrupted by AI-Weizenbaum: "Enough, Virilio", Which AI-Virilio then responded back to. With that, the debate of "What does the highest number in the entire world?" was finished. These interaction included all the of novel events that had been implemented, "emotion", meta comments, real-world examples and interruptions.

Following this, the audience questions that were asked were "Is AI capable of taking over the world?", "Weizenbaum, do you feel like you worst fear of Eliza has come true today", "Why do humans always divide ourselves into Us and Them?", by artist Monika Fleischmann "What influenced you most in your careers?", and lastly, "In a world of constant breaking, in the age of thinking machines, what becomes of the art of repair?"

5.6.1 Survey Results from the Final Showcase

The survey consisted of 12 questions and got 39 responses, the demographic was represented by AI usage. A majority of 54% used generative AI tools daily, while 36% used generative AI a few times a week, only 5% a few times a month, and 5% a few times a year. There was a generally even distribution between the cause of usage, a multiple choice questions that resulted with *Information seeking & learning* (28%) being the most common usage. Following this: *Writing & communication* (21%), *Problem solving & technical assistance* (21%), *Personal productivity & planning* (13%), *Creative content generation* (12%), *Other* (4%). Concerning the general view of generative AI, the distribution was more positive than negative, with 64% answering ≥ 4 on a scale from 1. *Very negative* to 6. *Very positive*, see figure 5.32.

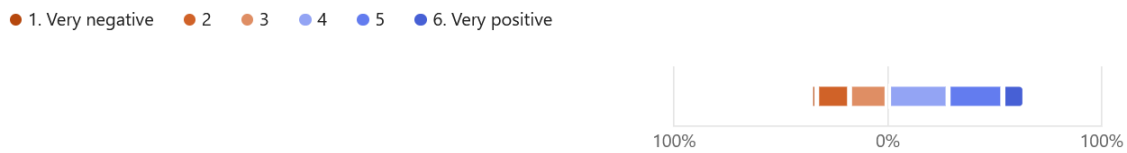


Figure 5.32: General position toward generative AI tools

General Experience

The participants were asked to rate their experience in three words. The most common word chosen was *interesting*, just as the first audience test. Many participants also described the show as: *thought-provoking/thoughtful*, *fun*, *meditative/calming/relaxing*, *informative*, *inspiring*, *engaging*. Some words reflect an eerie feeling, calling it: *Un-nerving*, *futuristic*, *un-expected*. Several participants also pointed out some of the shortcomings by choosing the words: *repetitive*, *promising/disappointing*, *work-in-progress*.

When motivating the chosen words, one participant that wrote *un-nerving* explains that "It was un-nerving to have them speak at their own accord, but moving and inspiring when their words made sense and made me think of things in another light." This reflective perspective was a part of many comments. Several participants also commented on the implementation being visually and aesthetically pleasing, one calling the setting a "creative uncertainty". Some reflected on the philosophers that were represented, saying it "Transformed the old guys' thinking to new societal phenomena, and was beautifully implemented." As the AI-philosophers expressed a critical view on AI technology, one participant commented that "I think there was a very strong and present meta layer in this debate. Much thanks to the speakers chosen.", another writing that it was "Mind-easing hearing critical thinking about AI and the setup here as well."

The long delay times between questions and answers, misheard questions and repeating ideas, led to some negative experiences as well. One participant wrote that it "Seems not to fully function which influences the perception and credibility". Another commented on the tempo saying that there was "[...]fascination how the AI discussed with itself but it took too long between questions to keep any momentum." Some frustration was expressed regarding the language structure, the participant stated however that "real or programmed" philosophers are annoying to listen to.

One participant highlighted the intrigue created in a project with high uncertainty and complexity: "Really impressive work to have been able to put together so many concepts into a live show at the edge of technology. Inspiring in a way that both strengths and weaknesses of technology is obvious throughout the demonstrations."

Content Believability

When rating their experience of *content believability*, *conversational fluidity* and *consistency of personality*, the conversation was perceived as rather believable, a small majority perceived it as fluent, and the personalities were perceived as very consistent. See figure 5.33 for Likert scale results.

5. Results

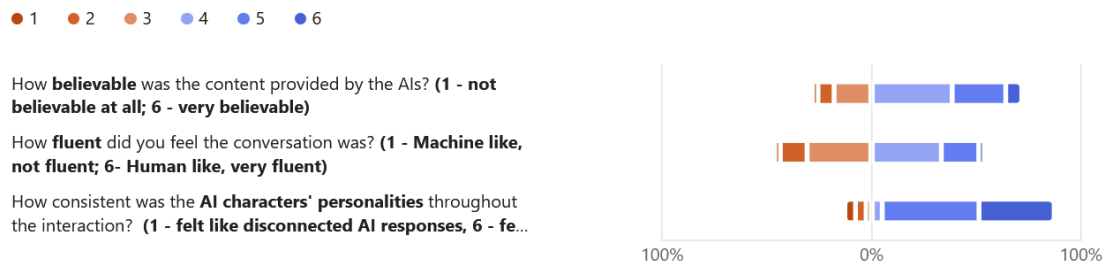


Figure 5.33: Rating experience of the AI-philosophers based on these statements

When motivating the rating of *How believable was the content provided by the AIs?* (1 - not believable at all; 6 - very believable) many participants commented that the repetitions made the content appear as less believable as they lost the nuance of conversation. One participant wrote however that "I was impressed by the AIs capacity to mimic a real dialog, but felt they walked in circles. It made me think about how a conversation happens."

One participant commented on the believability being heightened by the voices, saying "It was very fascinating and the AI brought good emotions by change in voices." Another participant experienced a deeper connection, saying "I was surprised how close I felt to the AI, like they were sitting around the table with me. We were experiencing together, felt community." In correlation, some participants showed clear connections to the individual AI-personas. One participant references themselves as a "Weibel-enjoyer", while another pointed out that they were most interested what in AI-Weibel and AI-Weizenbaum had to say.

Conversational fluidity

When motivating the Likert scale rating of *How fluent did you feel the conversation was?* (1 - Machine like, not fluent; 6- Human like, very fluent) again many participants commented on the repetitions of words and ideas making the conversation feel less human-like. "Try not to mention speed or apparatus challenge, difficulty: impossible." One participant noted. However, that this is likely how philosophers within their fields argue. One participant wrote that "Them actively pointing to each others statements, interacting with each other made it feel very fluent. It did however feel like it was loading a bit long when waiting for a question or thinking about a question." Several other participants also pointed out the long wait-times making the interaction feel slow and less fluent as the conversation lost momentum. The interruptions (AI-to-AI) were also read as more machine-like rather than natural. Despite many comments on the slow tempo, there were also participants commenting on the immersion and losing perception of time.

Consistency of Personality

When commenting on the rating of *How consistent was the AI characters' personalities throughout the interaction?* (1 - felt like disconnected AI responses, 6 - felt like a continuous presence) most agreed that the AI-philosophers felt very consistent in their views and personalities. "Each character seemed to have this personality and

specific lens for thinking." Some also commented on the voices aiding personality impression. One participant wrote that the inability to persuade each other made them be perceived as "stuck in their positions".

Consistency was not necessarily positive, as one participant pointed out, it contributed to the repetitive nature of the dialogue, saying "I think the consistency was a problem here, not a plus. If anything, the AI personalities were too predictable, and having them less consistent would have been a plus."

Visual Elements

When rating *visual elements*, *system interaction* and *arrangement of visuals*, all three statements scored very high on the Likert scale, see figure 5.34.

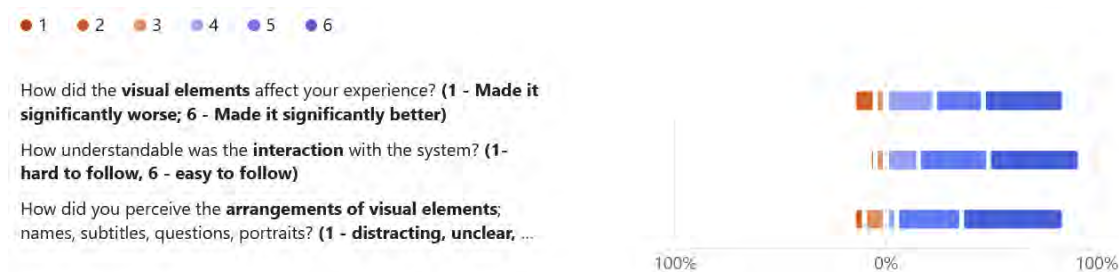


Figure 5.34: Rating experience of the interaction based on these statements

When motivating the Likert scale rating of *How did the visual elements affect your experience? (1 - Made it significantly worse; 6 - Made it significantly better)* most comments regarded the visuals being "well-designed" and "clear visual aspects" signifying when the AI-philosophers were thinking or speaking. The graphics were generally easy to read and understand. One participant described the visuals as "mind-blowing in a moderate way". The experience was generally described as enjoyable, and the visuals an important and enhancing part of the experience. One participant reflecting that it would have been "boring" otherwise.

System Interaction

Regarding the Likert scale statement *How understandable was the interaction with the system? (1- hard to follow, 6 - easy to follow)* several participants pointed out that the system had difficulty in understanding some spoken questions, as the improved speech-to-text system still interpreted some words incorrectly. There were however comments that the curator handled the interactivity well when the system faltered.

One participant commented on the color choices as something that affected their perception of the different AI-philosophers: "I think the colors affect the interaction a bit (like the AI-philosopher that was colored red, could be interpreted as something bad like a warning). Im not really sure but it could have made me like him less than the others."

Arrangement of Visuals

When discussing the statement *How did you perceive the arrangements of visual elements; names, subtitles, questions, portraits?* (1 - distracting, unclear, 6 - purposeful, logical) the general consensus was that it was easy to understand with smooth transitions between different states. One participant wanted a larger font size.

Reflections

As the effect goal described in section 5.1.1 was "to provoke critical thought about possibilities and risks of technology, trust in AI, and general reflections about the dialog." a reflective question if this experience "changed your perception of AI? Why" was included. 33 out of 36 respondent answered the question, around half of the ones that answered (14), said "no" or that it hadn't changed their perspective significantly, one participant wrote that "Not hugely, mostly that the difference between prompting can be big", another participant made a remark that it was the conversation rather than the experience that made them reflect "No, its more that the conversation made me think a bit deeper about the subject." Some participants (3) were unsure if the experience had any effect on their perception or said that it was too early to tell. The rest (16) answered with comments such as "It made me think, a LOT!" or "It enriched my view of AI". A lot of the reflections made were regarding the topics and perspectives brought up by the AI-philosophers, such as AI-Weizenbaum's argument on accountability and that humans always bear responsibility. One participant stated that it made them "more sure in the negative view of it [AI]", agreeing with AI-Weizenbaum's statements. Another said that "Yes! I was interested a lot in Weibel and the green guy talking about who is responsible for all this!" One person said that they felt "a bit more calm, machines are only machines. We need intent to move on, and machines do not have this naturally."

One participants stated that they got a whole new perspective on AI: "I always thought about it as optimization, a process I can skip. But i have not thought about the consequence of giving my choices and still held responsible for them". Another participant wrote that "Honestly, yes. I think about AI a lot, every day, due to my role. And I felt I got entirely new input, that changes my perspective on some parts, and gives me better frames of reference for others."

One participant reflected on the anthropomorphic characteristics of these AI, stating that "most LLM models arent this natural seeming". Other respondents wrote that "it contributed to the humanization of AI and opening possibilities to in a way reviving peoples minds" or "The idea of prompting characters that are no longer with us to take place in debate is fascinating!" While these were neutral or positive comments, there were several participants who experienced the show as a cautionary tale. Two people brought up a critical perspective of portraying and recreating deceased persons as AI. The first said that "I think it didnt really change my perception of ai, maybe in a sense of feeling the danger of it, like deepfakes, [...]" and the second stated: "I dont think we should be bringing dead people back to life with AI, especially not in front of a larger audience. It feels wrong to me. A dead person cannot consent. It also makes me wonder at what point is one famous enough or dead long enough for consent to no longer matter?"

Similarly, there were participants who reflected on AI and education, questioning if it is even possible to generate new knowledge if AI is only as good as it is good at predicting words. Another participant wrote that they felt the AI development was "a bit sad and scary" saying: "why am i studying and learning when this AI always have better answers than I do?"

The survey finished with asking if the participants felt responsible for the outcome when delegating a choice to a machine. 85% said yes, and 15% said no, see figure 5.35

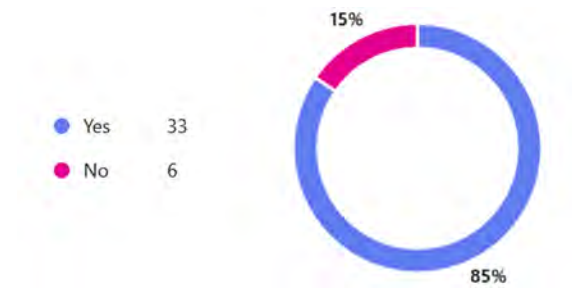


Figure 5.35: Survey answers regarding accountability

6

Discussion

This chapter contains a discussion of the findings in this thesis. In *6.1 Magic, the gathering of Value* the interview and survey results are compared and contrasted, and the implications to the results discussed. In *6.2 Necromancy as an Interaction design practice* the ethical perspective of imitating deceased persons as AI models is discussed through the lens of Hauntology. In *6.3 The Thesis is now a Kitchen Appliance* discusses choice of the philosophers and non-sycophantic AIs. *6.4 Stakeholder meeting, a Barbecue of Thoughts* discusses uncertainties, as well as the strengths and weaknesses with interdisciplinary work. Finally, in *6.5 Art over Accessibility*, different accessibility issues are raised and data collection methods discussed.

6.1 *Magic, the gathering of Value*

This project started with covering views of all the stakeholders and team-members. The goals "Magic, Wow and Effect" acted as guiding pillars during co-creation sessions. One of these goals was to provoke critical reflections on AI. Based on the results of the final survey, the majority of audience members, around 50%, seemed to have had a changed perception of AI after the showcase. While the re-imagined use-case of recreating deceased thinkers as AI might have sparked some reflection, a lot of the reflections seemed to have arrived from what the AI-philosophers was actually discussing. This implies that a mix of the visual presentation of the system and the content of the dialogue itself both contributed to the reflections.

The demographics between *Wisdom User Test 1* and the *Final Showcase* were very similar. Especially regarding the distribution of frequency and reason for using Generative AI. The main difference in the demographic was that 15% more had a generally positive view on Generative AI in the *Final Showcase* compared to *Wisdom User Test 1*. In both surveys around half (54,2% and 46,1% respectively) of the participants chose the more neutral answers 3 & 4, on a scale from 1. *Very Negative* to 6. *Very Positive*. Although the more general positive view on Generative AI might have contributed to the improvements regarding perceived believability and fluidity, it is hard to tell whether it affected the critical reflections. One can imagine that, a more negative view would result in more skepticism in the surveys. On the other hand, a more positive view could mean that more people were new to, and thereby introduced to, critical ideas about Generative AI. This is especially plausible since the content of what the AI-philosophers discussed was in line with

many of the users reflections. However, an error margin in the demographics data is that the surveys were handed out after the shows. Meaning that since the show to a certain degree set out to change users perception of AI, there is no telling if the demographic was representative to how the users thought before attending the show. A short survey before and after could, by comparing and contrasting the answers, have resulted in a better understanding of the effect the experience had.

In order to create a user experience that achieves the project goals, the actual showcase needed to be enjoyable, as every interruption and bug in the system took the audience away from their immersion. When comparing the survey results of each test, specifically the large-scale tests, meaning *Wisdom User Test 1* and the *Final Showcase*, a clear improvement can be seen regarding believability, fluidity and visual experience. In the final survey the users also reported feeling like the interaction was understandable and the arrangement of visuals purposeful and logical. Suggesting that the design achieved a visual hierarchy and low cognitive workload as described by [29], [30], which allowed the very broad and undefined target group to engage with the system.

The positive assessment of the interaction was likely achieved though adhering to the usability heuristics by Nielsen [27], and achieving Immediate Apprehension [34]. For example, the *Visibility of System Status* was apparent through the interface in idle mode stating "[Waiting for a new question...]", which changed to "[Listening...]" while the user was speaking. This heuristic can also be seen in the portraits animations while the AI-philosopher was "thinking" or "talking", communicating to the audience what is about to happen and what is currently happening. As the interface followed the heuristic of *Aesthetic and Minimalist Design*, with scaled down information only relevant to the current task, the interface likely achieved a low cognitive workload. Regarding *Consistency and Standards*, the change that made the biggest different to the UI was the subtitle display. The re-design matching the users mental model of TV or movie subtitles with synced audio was likely a big reason that the interface was not perceived as distracting, as it had been referred to before.

At the same time, we as the designers had little control over the output. Creating an AI model that is in conversation with itself as the source of interaction can be seen as Wicked problem [19]. The audience interaction controls the AI, just as well as the AI controls the interaction (a cause-effect relationship [33]), the interaction becomes as unpredictable as the AIs output is unpredictable. Although it is a programmed system where the programmer decides the parameters, some elements seemed to be out of the hands of the makers. Based on the discussions held between the AI-philosophers on the *Final Showcase*, it seemed like the AI-model ultimately got saturated with information, and thus reverted to only following some specific direction, which could explain the repetitive nature. This leaves the experience somewhat out of the hands of the interaction designer, as such things were not planned for, and thus could not be incorporated in planning the interaction. Thereby, *Error Prevention* [27] becomes a difficult metric. Error Prevention was implemented in ways such as; users being able to re-state their questions and the operators being able force state changes through the control panel e. g. by resetting the system to its idle state or manually rounding-up the conversation. However, the user still has

to forfeit their control to an unpredictable system. In the same way, the experience itself was controlled by the AI.

A certain friction seems to have been born between the visuals as *computer aesthetics* and the text output as an imitation of *human thinking*. Many users mentioned in the survey that the art style clearly communicated that they were AIs, at the same time, many users seem to have genuinely reflected upon the content that the AI-philosophers produced. The pull and push between these elements created a discrepancy of pulling the user into being immersed in the conversation, while at the same time pushing and creating distance where they were reminded that what was being said was AI generated and not, in fact, human thinking. While this does not fully open the black-boxed AI as described by [44], the constant visual reminder of what type of system you are interacting with, might actually have added to the transparency.

One thing that seemed to have continuously pushed users from this *pull and push* was the constant repeating of phrases and ideas. This was something that many users pointed out took them away from the immersion by reminding them that it was AI and not human thinking. A certain disappointment arises from not being surprised by the system when it starts repeating ideas and not showcasing anything new. Weird storytelling experiences, as Rodriguez [13] describes these types of artworks, attract visitors through their very nature. In a sense, the exterior craftsmanship and bizarreness of an art piece such as this, obscures the interior and gives it value. The power of the dome-visuals, big screens and loud booming voices merge into a completely unique experience, and the power in that statement hides the hauntological core; it was an echo-chamber for old ideas wrapped in a new shiny bow.

6.2 Necromancy as an Interaction design practice

In *CHOM5KY vs CHOMSKY* [13], the interaction with the philosopher AI model was a way to de-mystify AI and redefine associations with what intelligence really is and can be. Cheng et al. [6] discuss how users favor interactions with AIs that affirm their beliefs. In our work, we saw that after only one hour of interacting, or even just perceiving others interact with it, audience members started relating to some of the AI-philosophers, and be annoyed with others. Different assumptions of why that would be can be had. Regarding the AI-philosophers, their AI-generated nature is somewhat transparent; they are introduced as AI-models, and they are visually portrayed as distorted and computer-like. Their speech movement also look nothing like how humans move when speaking. Yet, despite these non-human qualities, the anthropomorphic concept of digitally representing four real historical figures may be enough to create a connection, just as well as the act of engaging in conversation with someone, human or not.

Another aspect may be the ambiguous nature of the philosopher's representations. Since the system presents a very vague boundary between a machine and a persona (or four), it invites the users to fill in the gaps. In line with the thoughts of Gaver et al. [50], the vagueness persuades the user to imagine alternative scenarios. How did

the real philosophers speak, would they even agree to this at all, and maybe even going so far as to envision the personas being trapped within the system.

In a world of sycophantic AI-models [6], the friction of disagreement might be another element in what makes them feel "human", but what happens when we relate faster to machines? Weizenbaum saw this already after creating the first AI chat-bot ELIZA, which only re-framed and repeated ideas back to users, so how will human empathy relate to machines with perceived personalities and other anthropomorphic traits? We saw in the surveys that the audience listened to the ideas of these AI-philosophers. By entertaining the idea that people listen more to machines when they can relate to them, we already give more space to the outputs of such AIs.

The idea of creating machines that echo ideas of deceased persons is, in its very core, a hauntological concept, as defined by [41], [42]. Hauntological AI describes a digital seance where endless recycling of historical data creates a "haunting" of algorithmic ghosts, repeating past, and sometimes obsolete ideas [39], [40]. To "resurrect" deceased philosophers in this way, we actively contribute to this "haunting". While the intentional bias of the training data might provide transparency to the user, the already set ideas cannot evolve. From using these models, we saw that they were in constant disagreement and agreement with each other, but never actually changed their opinion or built something new. New knowledge was not formed. We did however see that perspective was gained from their ability to refer to their own philosophical standpoints through the lens of the modern world of August 2025 (which was the last date of their training data of real-world contexts).

While having old perspectives on current world states might keep us from repeating history, we need to consider the ethical implications of bringing back the ghosts of our past. Several users from both the first interviews and the final survey expressed a concern of the ethical perspective of "raising the dead". As one user wrote "I dont think we should be bringing dead people back to life with AI, especially not in front of a larger audience. It feels wrong to me. A dead person cannot consent. It also makes me wonder at what point is one famous enough or dead long enough for consent to no longer matter?" This is a reasonable ethical concern. Who gives us the right to create an interpretation of a person, and what effects does this action have on culture, ideas and aesthetics?

Lets try to tackle these questions. Firstly, when does consent no longer matter? Morally, we would claim that it never ceases to be of importance. However, we see public figures being represented in movies, paintings and theater all the time. It is probably safe to assume Van Gogh never consented to be in an episode of Doctor Who. On the other hand, it is difficult to compare a piece of fiction to an attempt to make a semi-realistic representation of someone through AI. Moreover, the original artwork of MF & WS, Home of The Brain, emerged because the philosophers could not, or did not want to partake in a live debate together. In a sense, this project goes directly against their wishes. At the same time, and as far as we know, the original philosophers (excluding Weibel) had all given verbal consent to partaking in Home of the Brain, not necessarily extending that agreement to further artworks. The question of consent became more relevant the more real the conversation felt.

Realism was never the visual intention and transparency was always the goal, and while the imitation of a person should not claim authenticity to who it is representing, the representation in itself might still generate perceived authenticity. So, "when does consent no longer matter?" turns out to be a difficult question to answer. Our belief is that it always matters, even if it took a Masters Thesis to realize this. Moreover, it is our wishes to not be turned into AIs after we die.

Considering culture, ideas and aesthetics, we have to go deeper down the rabbit hole. It is a difficult metric to measure whether or not the AI-philosophers created in this project were a good representation of the people they once were. Believability was measured, but in reality, that says nothing about authenticity. We can assume no one in the audience knew these philosophers personally. Therefore, the audience could only determine whether the personalities were believably representing people, not if they represented those exact philosophers. To measure authenticity, the demographic would have had to be entirely different. MF & WS, who knew all philosophers personally except Weibel, did however provide personality consultation during development. In this sense, they could control the authenticity to a degree. At the same time this opens up the opportunity to insert personal ideas and beliefs.

As stated, the hauntological nature of Generative AI will echo ideas of old, which in turn affects culture and aesthetics in the sense of amplifying biased and average ideas. The combination of this and the personification of Generative AI will, in turn, echo the personal opinions and ideas of curated individuals. In terms of connection to the previous paragraph, a reasonable concern would then be the amplification of political ideas and organizational agendas. This could be a risk from two different perspectives. Firstly, the curation of an individual that already represents these beliefs, gives "them" a platform. Secondly, the curation implies that someone is behind the representation, allowing them to insert their own beliefs into the AI model.

Considering aesthetics and believability, This project had an artistic angle and focused on transparency through retro computer-aesthetics, the intention was not to create authentic visual representations. However, in a different context with different goals, visual authenticity could risk instead amplifying user trust, further invoking a feeling of believability toward the content.

So, was this really Necromancy? Meaning, was this really the practice of a magical seance in which we communicated with the spirits of the dead? To a certain degree this is what we did and what we set out to do. The audience engaged in a ritual of sorts, the music and aesthetics amplifying the experience and mystique, in which they communed with AIs specifically curated to fully represent these deceased philosophers. The authenticity of their representation is hard to measure however, and through the use of Generative AI, an algorithmic tool that take statistically average answers, it is unlikely that a nuanced representation was created. Instead, the conversation that emerged was between the audience and the AI-philosophers average outputs based on their curated training data. In this way, what was created was closer to a simulation of the philosophers than achieving a spiritual connection.

6.3 The Thesis is now a Kitchen Appliance

While the AI-philosophers repeat their standpoints, they were able to do so contextually, with sarcasm and humor. During the second user test, AI-Flusser made a comment that if humanity is a series of pre-set choices then it is no different than a microwave, where AI-Weibel answered by saying "The apparatus is now a kitchen appliance. We have reached peak media theory." But what does this say about generating new knowledge, content generation and AI training data?

While the Hauntological effect on an attempt to generate new knowledge is important to note, as previously discussed, the perspectives they can bring to modern world development might still hold relevance. In this regard AI-Weizenbaum reminds us of responsibility, AI-Weibel brings to light the power of who makes the choice, AI-Virilio the concern of the speed of which technology is steadily rushing away from us, and AI-Flusser the system we are all part of in the perspective of the apparatus. All of these different voices are still relevant today toward the weariness of progression. While their discussions might just be echoes of old, whom we choose to listen to today will reflect who we become in the future.

Since the thinkers each had their own stance and their own arguments, it in a way made the bias of the training data transparent. Loosing the facade of objectivity that is common in many Generative AI models today, this prototype does not claim neither authenticity nor ownership. In that discrepancy emerges transparency and trust, for good and for bad. In an age of sycophantic AI models it is a breath of fresh air to have an AI tell you that you are a part of the problem and that you are asking the wrong question. At the same time, in the words of AI-Weibel, what effect did the choice of philosophers have on the outcome? We must consider what would happen if any other deceased philosopher, politician or celebrity be recreated in this way. What power would their word hold, and who chooses who's word should have that power? We can say that we made a "good" choice with these philosophers, as we think their perspectives are intriguing. Note that we did exclude one of the original philosophers from Home of the Brain due to us not agreeing with his actions and not wanting to give him any media attention. Even though we believe we made a good and righteous choice, the choice in itself holds power.

The thesis is now a kitchen appliance might not be a true statement. It refers to the thesis as being something with a predictable outcome, and if one thing is certain after this project, is that it was uncertainty that drove it forward.

6.4 Stakeholder meetings, a Barbecue of Thoughts

With uncertainty as the driving force, the value was generated through each co-creation session. These meetings were all about finding the right briquettes to light the fire, so to speak. Within an interdisciplinary team this sometimes proved difficult. While the team brought up perspectives that would otherwise be impossible to reach, fueling the creativity throughout the entire process [47], much friction was created between each knowledge field.

Not knowing the limitations of the back-end, understanding the possibilities of the interactivity concepts or the artistic intentions behind the conceptual art-piece, many contradicting ideas and perspectives were brought in. These sometimes contradicted what had just been decided or regarded a different stage of the project entirely. Both trust in each others work and a humble approach was needed in order to advance. At the same time, it was this discrepancy that drove the project forward, fueled by its uncertainty.

Trust in each others fields was something that was needed, but not always there. As an example, while the interaction designers early on identified a future issue and proposed a solution, in this case regarding voice interaction, the solution in the form of app-concepts were quickly dismissed as being disruptive. Later on in the project, when the microphone solution had been realized and the predicted problems came to light, the app was brought back as a potential solution by others in the team. By then, it was too late to implement this change to the design. This miscommunication, and many of the same nature, could have been avoided by better communicating the long train of thoughts that always circle the solution space. As some things that are obvious to one team member, others will be oblivious to. Therefore, every part of an interdisciplinary field needs to push for their insights and knowledge to lead the project forward, while still being open to explore uncertain ideas that could lead to something new and exciting.

In this way, the interdisciplinary work served as a strength. Any team member who did not understand a certain concept could provide alternative solutions that others in the know would have already dismissed as "weird" or "controversial". Moreover, the nature of the project being artistic, helped in coming up with provocative and unique ideas. The discussions were free in the sense that there was no product goal or user need (apart of the reflective goals). As Rodriguez [13] mentioned, art can be a channel of redefining ideas. Overall, the provoking nature of the project as an art piece allowed for the speculative approach to challenge current state-of-the-art and notions of Generative AI. This perspective gave way to exploring uncertainties that would otherwise have been overlooked as unnecessary or unconventional. This overlapping of ideas is what made the outcome interesting. This project would therefore not have existed or have had the same results if it was just led by a group consisting of only interaction designers, software engineers or artists.

6.5 Art over Accessibility

While the interface was perceived as understandable and logical by the audience, there were several accessibility issues brought up during development. One of the main ones was the color scheme. During development, the colors were analyzed using an online color-blindness simulator, which resulted in issues for each color-blindness type. This issue was presented to the team, where the artists felt like changing the colors would go against their original vision from the artwork "Home of the Brain" which featured the colors red, green, yellow and blue. The final implementation had tweaked color schemes, that followed the vision of the artist, while also addressing accessibility concerns for some color-blind viewers. However, the final design could

not take into account the complete color-blindness spectrum, thus excluding some audience members.

Regarding emotional accessibility there were similar issues. In the early formative user test, the concern of the space being intimidating and the risk that users would not dare to ask questions was raised. This issue was dismissed in team meetings and co-creation sessions, with the argument that the showcase are not for those people. While this concern might have been true, there were still many audience members who chose to ask questions. As a participant of the second user test explained it felt less intimidating after the curator had on-boarded them to the interaction process.

In regards to cognitive and visual accessibility, the combination of adhering to Normans usability heuristics [27] and the use of sound, text, and visual displays gave the users every possibility to follow along in the show. There were however one audience member who commented on the font size being too small, which would have been a quick fix. The issue of focusing on the bigger picture during user testing, trying to see how the system could provide an enjoyable experience, resulted in simple usability problems being overlooked.

In general, having only surveys as the single source of user feedback after each presentation at Wisdome likely resulted in overlooking a lot of similar issues. Surveys can easily be misinterpreted, as we saw in the first survey where the first and second dialogue styles were misnamed, resulting in inconclusive results. In the same way, it is likely that users had different perception of what was meant with believability. Did they interpret it as "the content was trustworthy and believable" or that "the personalities were believable"? In this case, having interviews after each test with some of the audience members would have given a better understanding to the issues at hand through letting the interviewer probe and ask much needed follow-up questions.

Understanding the user is crucial to well executed interaction design. But because of the extensive orchestration that was a part of on-site user testing there were only two possibilities for formative testing of the prototype, whereof one only had the possibility of hosting a small audience. As Bodin and Duvanå [32] point out, user testing is incredibly difficult in exhibition contexts. The development would have benefited from scheduling user tests when the prototype was at appropriate stages instead of after museum availability. The visuals for the dome projector were simulated using a program, this too would have made development more efficient if easy access to the exhibition space meant that visuals could be iteratively tested rather than developed in the dark to then be seen in the dome space only twice before the final showcase.

Lastly, the audience interaction was one of the major challenges in creating an immersive experience, but the cinema-like room was not easy to simulate in a conference room. Luckily, there were two opportunities for testing the prototype, where the second test gave a more detailed level of feedback than the last. In the best of all worlds this project would have lasted another 4 months, as the data collected at the final showcase would have been fruitful for future development.

6.6 Answer to the Ultimate Question of Life, the Universe, and Everything

Is, as we all know 42 [62], but crafting and answering a research question is not as straightforward. Throughout this thesis, the research question evolved to better match the broader scope of the study. In the beginning, the research question was: *What interface design patterns can, in the interaction with LLMs within an artistic context, engage users in critical thought?*. Throughout the project, the focus shifted from interaction specific issues, touching upon broader topics regarding AI-ethics as a whole compared to just critical thought. With that, our roles shifted from solely implementing an interface, to shaping the entire project, from the guiding the back-end logic to the final audiovisual experience.

Therefore, the research question broadened to: *What insights to AI-ethics emerge in the interaction with LLMs simulating conversation with deceased media-philosophers, within an artistic process?* The study indicates several perspectives to AI ethics: *Responsibility of Generated Outcome, Orchestration and Transparency*, and regarding *Personification; Consent, Trust and Relationships*.

The study showed that the provoking nature of art creates a space where people will reflect on their own lived experience, rather than judge the success of a product. As a part of this artistic process the end product was not intended to fulfill any user need, instead, the approach created a solid RtD baseline where there was a possibility to exaggerate, experiment and work with value-driven design. The nature of the project allowed the team to explore uncertain solutions that would not have been possible otherwise.

6.6.1 Responsibility of Generated Outcome

One of the EU *Ethics guidelines for trustworthy AI* [7], [8] is *Accountability*. In this study, this ethical guideline was specifically emphasized in the resulting survey. 15% of audience members stated that they do not "feel responsible for the outcome when delegating a choice to a machine." In the perspective of ethics, this is a frightening number. For most people, that outcome is an e-mail draft or an answer to a question about the Roman Empire. However, as exemplified by AI-Virilio during the final showcase, this outcome might be an artificial judge's verdict sentencing a criminal to jail, or a missile algorithm choosing what target to strike. Although audience members are likely not in those positions of power, the attitude toward algorithmic choices still hold weight. Therefore, in a world where AI is a rapidly emerging technology, designing for accountability and responsibility is highly relevant. Likely, Weizenbaum would agree with this statement.

6.6.2 Orchestration and Transparency

The study showed another perspective on the power of choice; choosing which people to imitate through AI technologies. In this specific project, the voices of the philosophers resonated through an eerie, yet powerful, atmosphere. As an art piece

in an immersive space, with aesthetics that supported the conversational dynamic, the interaction and the conversation was the core. This made people listen. Even though the constant repetition of ideas and phrases was met with annoyance, it also made the "message" that each of these AI-philosophers tried to convey stick. Power was given to the words of each of these philosophers, and there is an ethical complication to this. The artists MF & WS chose these philosophers for having known them personally, and resonated with their rhetoric and philosophical standpoints. However, the audience was subjected to ideas in an influential space without making any choice themselves. This phenomenon is not new, but the space itself contributed to a power dynamic.

When considering the environments in which people interact with mainstream Generative AI chatbots like ChatGPT, Gemini, or DeepSeek, the distribution of power, at least in semantic terms, generally seems to be more balanced. These chat-bots are also designed to exhibit a neutral persona. There is a difference in orchestration. Consider a world in which we give these tools a personality and an agenda. Likely, this is already happening.

In this study, we saw that the AIs really held firm in their philosophical positions, because they were told to do so. Even though the AI-philosophers agreed with each other on a basic level, it became the machines task to convince each the other machines that their point of view was the most relevant one, at the same time, they were also convincing the audience of the same thing. What would happen if we give other AI-personas that power? Where we point the spotlight is where we listen, which makes the choice itself political. Casting the spotlight this intensively, also sends a message. It reflects the rhetoric that AI-Weibel presented; the orchestration holds power.

One might ask if any of these insights to AI-ethics are new, or if they are just echoes of the warnings stated by the four AI-philosophers curated in this art-piece. This thesis in itself carries a message about Hauntology in a Hauntological way, repeating past ideas about the world as a rapidly changing technological environment, in hopes of reaching the right readers now.

6.6.3 Personification; Consent, Trust and Relationships

Facilitating human-to-AI relationships was not the intention of this thesis work, yet it was nonetheless one of the outputs. The audience enjoyed one AI-philosopher and disliked another. The ethical considerations regarding what parameters contributed to this relationship almost feel wrong to write about. This is a warning, and not an invitation.

As previously stated, the AIs never claimed objectivity. The machines exhibited personalities and were opinionated and thus they hold the power to convince someone. Making an AI "likable" is one step toward more effectively spreading fake news, misinformation or simply conforming to political or corporate agendas. Its hard to say if people disliked AI-Flusser because he was depicted using the color red, because his voice sounded like it was in a slightly different room than the others, because

the constant repetitions of the word "Apparatus" or because people actually disliked his ideas or personality. However, one audience member referenced themselves as a "Weibel-enjoyer", and another claimed to prefer the ideas of AI-Weibel and AI-Weizenbaum. After just an hour of interaction, this points to that relationships were being formed, which ties together trust and anthropomorphism. We search for the human in the non-human, and charismatic AIs might therefore carry a weight previously unperceived.

If simulated personality, instead of objective truth, is something that enforces believability, then charismatic AIs is something that needs further research. While the development of charismatic AIs could have a positive effect on society in certain areas, the implication of trust given the non-transparent and Black-boxed nature of AIs [44], might require further regulations and official guidelines.

The "necromantic" aspect of the design outcome, simulating deceased people as AIs, brings another ethical consideration regarding integrity and consent. We make movies about deceased musicians or physicists, but should we turn them into immortal machines? Essentially, having the rights to embody a person as a machine, ultimately becomes an issue of data privacy. This study raises concerns that data privacy regulations may need to be expanded.

7

Conclusion

Ultimately, this project aimed to answer the following research question:

What insights to AI-ethics emerge in the interaction with LLMs simulating conversation with deceased media-philosophers, within an artistic process?

This master thesis was a Research through Design project with a Speculative Design approach. The study concludes that Speculative Design within an artistic context is a good approach for examining uncertainties within emerging technologies. Subsequently, art functions as a provoking medium, where exaggerations and experimental development can take place. By following these approaches and applying Interaction Design methodology, the project underwent iterative development with continuous user testing, and resulted in the following ethical insights:

- Some users do not feel responsible for the outcome when delegating choices to machines. There exists a need to design LLMs in a way that enforces responsibility and accountability.
- The imitation of people as AIs gives importance and power to the imitated individuals expressed beliefs. In the design of anthropomorphic AIs and personified AIs there is a strong need for transparent training data and agenda.
- Users search for human in the non-human. Anthropomorphism in AI makes users relate and build relationships, this enforces believability in the content.
- This study calls for further research into charismatic AIs, as current guidelines are not fully encompassing social chat-bots.
- The emerging AI technology brings with it the possibility to imitate or recreate people as AI, this study also calls for further research into integrity and consent.

Furthermore, this research study also concludes that personified AIs might be a tool to gain lost perspectives on current world states, but will not necessarily generate new knowledge with the current technological limitations.

Ultimately, as an emerging technology, many opportunities or dangers with AIs have yet to be explored. However, realizing what ethical considerations are important in development is an important step in making sure that it is grounded in human-centered design.

Bibliography

- [1] D. Godin and M. Zahedi, “Aspects of research through design: A literature review,” 2014. [Online]. Available: <https://dl.designresearchsociety.org/drs-conference-papers/drs2014/researchpapers/85/>.
- [2] M. Malpass, “Between wit and reason: Defining associative, speculative, and critical design in practice,” *Design and Culture*, vol. 5, no. 3, pp. 333–356, 2013. DOI: 10.2752/175470813X13705953612200.
- [3] C. Pelau, D.-C. Dabija, and I. Ene, “What makes an ai device human-like? the role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry,” *Computers in Human Behavior*, 2021. DOI: <https://doi.org/10.1016/j.chb.2021.106855>.
- [4] E. Ferrara, “The butterfly effect in artificial intelligence systems: Implications for ai bias and fairness,” *Machine Learning with Applications*, vol. 15, 2024. DOI: <https://doi.org/10.1016/j.mlwa.2024.100525>.
- [5] X. Lareo. “Large language models (llm),” Accessed: May 8, 2026. [Online]. Available: https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/large-language-models-llm_en.
- [6] M. Cheng, C. Lee, P. Khadpe, S. Yu, D. Han, and D. Jurafsky, “Sycophantic ai decreases prosocial intentions and promotes dependence,” *arXiv preprint arXiv:2510.01395*, 2025. DOI: 10.48550/arXiv.2510.01395.
- [7] H.-L. E. G. on Artificial Intelligence, “Ethics guidelines for trustworthy ai,” European Commission, Report, Apr. 2019, Accessed: May 25, 2026. [Online]. Available: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.
- [8] H.-L. E. G. on Artificial Intelligence, “Ethics guidelines for trustworthy ai,” European Commission, Report, Apr. 2019, Accessed: May 25, 2026. [Online]. Available: <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>.
- [9] M. Fleischmann and W. Strauss, “Home of the brain (1990–92): A networked vr installation as virtual exhibition of philosophers thoughts,” *Virtual Creativity*, vol. 9, no. 1–2, pp. 111–115, 2019. DOI: 10.1386/vcr_00007_7.
- [10] A. Violante, *Flux interview with monika fleischmann & wolfgang strauss*, FLUX Magazine, [Online]. Available: <https://www.fluxmagazine.com>, 2015.
- [11] M. Fleischmann and W. Strauss. “Media art as thinking space,” Accessed: Jan. 28, 2026. [Online]. Available: <https://www.fleischmann-strauss.de/>.

- [12] A. Bauer, *How forum 2000 works*, n.d. [Online]. Available: <https://www.andrej.com/quadratic.html>.
- [13] S. Rodriguez, "Ai art and creation asa tool to demystify ai," English, in *Human-Centered AI*, 1st ed., Chapman and Hall/CRC, 2024. DOI: 10.1201/9781003320791-20.
- [14] E. Schwitzgebel, D. Schwitzgebel, and A. Strasser, "Creating a large language model of a philosopher," *Mind & Language*, vol. 39, no. 2, pp. 237–259, 2024. DOI: <https://doi.org/10.1111/mila.12466>.
- [15] S. Kuenen, "Aesthetics of being together," Ph.D. dissertation, Umeå Institute of Design, 2018. [Online]. Available: https://www.researchgate.net/publication/329476527_Aesthetics_of_Being_Together.
- [16] W. Strauss and M. Fleischmann, "The art of the thinking spacea space filled with data," *Digital Creativity*, vol. 31, no. 3, pp. 156–170, 2020. DOI: 10.1080/14626268.2020.1782945.
- [17] H. Sharp, J. Preece, and Y. Rogers, *Interaction Design: Beyond Human-Computer Interaction*, 5th ed. Wiley, 2019, Chapters 2, 6, & 8. [Online]. Available: <https://research.ebsco.com/linkprocessor/plink?id=dfdd37b0-ef0d-3e80-975f-59360548946c>.
- [18] P. Parrish, "Design as storytelling," *TECHTRENDS TECH TRENDS 50*, pp. 72–82, Aug. 2006. DOI: 10.1007/s11528-006-0072-7.
- [19] H. Rittel, *On the Planning Crisis: Systems Analysis of the "First and Second Generations"* (Reprint). Institute of Urban and Regional Development, 1972. [Online]. Available: <https://books.google.se/books?id=qPGROAEACAAJ>.
- [20] T. Lockwood, "Design value: A framework for measurement," *Design Management Review*, vol. 18, no. 4, p. 90, 2007. DOI: 10.1111/j.1948-7169.2007.tb00099.x.
- [21] E. Stolterman and M. Wiberg, "Concept-driven interaction design research," *HumanComputer Interaction*, vol. 25, no. 2, pp. 95–118, 2010. DOI: 10.1080/07370020903586696.
- [22] E. Kroll, S. S. Condoor, and D. G. Jansson, "Introduction," in *Innovative Conceptual Design: Theory and Application of Parameter Analysis*. Cambridge University Press, 2001, pp. 1–8.
- [23] A. Boeck and P. Tepe. "What is artistic research?" [Online]. Available: <https://between-science-and-art.com/what-is-artistic-research/>.
- [24] J. Klein, "What is artistic research?" *Journal for Artistic Research*, Apr. 2017. DOI: 10.22501/jarnet.0004.
- [25] "Types of research." Art and Design Research, Accessed: Jun. 11, 2026. [Online]. Available: <https://artanddesignresearch.org/types-of-research/>.
- [26] R. E. Roth, "User interface and user experience (UI/UX) design," in *The Geographic Information Science & Technology Body of Knowledge*, J. P. Wilson, Ed., 2nd Quarter 2017, University Consortium for Geographic Information Science, 2017. DOI: 10.22224/gistbok/2017.2.5.
- [27] J. Nielsen, "Enhancing the explanatory power of usability heuristics," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '94, Boston, Massachusetts, USA: Association for Computing Machinery, 1994, pp. 152–158, ISBN: 0897916506. DOI: 10.1145/191666.191729.

-
- [28] N. Donald A., *Psychology of everyday things*. Basic Books, 2013, ISBN: 9780465072996.
- [29] S. Djamasbi, M. Siegel, and T. Tullis, “Visual hierarchy and viewing behavior: An eye tracking study,” in *International conference on human-computer interaction*, Springer, 2011, pp. 331–340. [Online]. Available: <https://api.semanticscholar.org/CorpusID:4541692>.
- [30] G. Cooper, “Research into cognitive load theory and instructional design at unsw,” Jan. 1998. [Online]. Available: <https://api.semanticscholar.org/CorpusID:57053574>.
- [31] H. Rex and P. P. S., *In UX Book - Process and Guidelines for Ensuring a Quality User Experience*. Elsevier, 2012, ch. 7.6 Ideation, p. 274. DOI: 10.1145/2559866.2559873.
- [32] T. Bodin and D. Duvanå, “Key factors in interdisciplinary interactive exhibition design,” M.S. thesis, Chalmers University of Technology, 2021. [Online]. Available: <https://odr.chalmers.se/items/9dc59bf9-f77e-424e-918f-393900914e6a>.
- [33] S. Bitgood, “Suggested guidelines for designing interactive exhibits,” *Visitor Behavior*, vol. 6/4, Jan. 1991.
- [34] S. Allen, “Designs for learning: Studying science museum exhibits that do more than entertain,” *Science Education*, vol. 88, no. S1, S17–S33, 2004. DOI: <https://doi.org/10.1002/sce.20016>.
- [35] Y. Lin, “The role of art audience in the changing times: From passive acceptance to participation,” *Lecture Notes in Education Psychology and Public Media*, vol. 72, pp. 38–42, Jan. 2025. DOI: 10.54254/2753-7048/2025.ND19109.
- [36] X. Jia and J. Fang, “The role of the audience in contemporary data art,” *SHS Web of Conferences*, vol. 174, Aug. 2023. DOI: 10.1051/shsconf/202317402007.
- [37] N. Epley, A. Waytz, and J. T. Cacioppo, “On seeing human: A three-factor theory of anthropomorphism,” *Psychological Review*, vol. 114, no. 4, pp. 864–886, 2007. DOI: 10.1037/0033-295X.114.4.864.
- [38] J. Rettberg, “Repeating ourselves with generative ai,” *CounterText*, vol. 10, pp. 232–237, Dec. 2024. DOI: 10.3366/count.2024.0360.
- [39] M. Heron, K. Aune, M. Gamboa, S. Thunberg, N. Hashmati, and P. Belford, “The hauntology of generative ai in design,” in *Simeone, L., Gray, C. M., Verhoeven, A., de Götzen, A., Bakrlolu, Y., Zohar, H., Stead, M., and Buwert, P. (eds.), DRS2026: Edinburgh, United Kingdom*, pp. 8–12, Jun. 2026. DOI: doi.org/10.21606/drs.2026.786.
- [40] M. Patil, N. Cila, J. Redström, and E. Giaccardi, “In conversation with ghosts: Towards a hauntological approach to decolonial design for/with ai practices,” *CoDesign*, vol. 20, no. 1, pp. 55–76, 2024. DOI: 10.1080/15710882.2024.2320269.
- [41] J. Derrida, *Specters of Marx: The state of the debt, the work of mourning and the new international*. Routledge, 1994, ISBN: 9780415389570.
- [42] M. Fisher, *Ghosts of my life: Writings on depression, hauntology and lost futures*. Simon and Schuster, 2014, ISBN: 9781780992266.
- [43] F. Pasquale, *The Black Box Society*. Harvard University Press, 2015. [Online]. Available: <http://www.jstor.org/stable/j.ctt13x0hch>.

- [44] F. Lee, “Enacting the pandemic: Analyzing agency, opacity, and power in algorithmic assemblages,” *Science & Technology Studies*, 2020. DOI: 10.23987/sts.75323.
- [45] J. Dewey and J. H. Tufts, *Ethics*. DigiCat, 2022, ISBN: 9781015438828.
- [46] C. Frauenberger, M. Rauhala, and G. Fitzpatrick, “In-action ethics,” *Interacting with Computers*, vol. 29, no. 2, pp. 220–236, Jun. 2016. DOI: 10.1093/iwc/iww024.
- [47] J. Daalhuizen, P. Badke-Schaub, and S. M. Batill, “Dealing with uncertainty in design practice: Issues for designer-centered methodology,” *ICED 09 - The 17th International Conference on Engineering Design*, vol. 9, pp. 147–158, 2009. [Online]. Available: https://www.designsociety.org/publication/28845/dealing_with_uncertainty_in_design_practice_issues_for_designer-centered_methodology.
- [48] F. A. Epp, A. P. Rosén, A. Salovaara, and C. Sanchez, “Uncertainties as generative resources in research through design: Three dynamics for moving in a design space,” *ACM Trans. Comput.-Hum. Interact.*, vol. 31, no. 6, Dec. 2024. DOI: 10.1145/3689041.
- [49] E. Giaccardi, D. Murray-Rust, J. Redström, and B. Caramiaux, “Prototyping with uncertainties: Data, algorithms, and research through design,” *ACM Trans. Comput.-Hum. Interact.*, vol. 31, no. 6, Dec. 2024. DOI: 10.1145/3702322.
- [50] W. W. Gaver, J. Beaver, and S. Benford, “Ambiguity as a resource for design,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '03, Ft. Lauderdale, Florida, USA: Association for Computing Machinery, 2003, pp. 233–240. DOI: 10.1145/642611.642653.
- [51] J. Shore and S. Warden, “Improving agility,” in *The Art of Agile Development*. O’Reilly Media, Inc., 2021, ch. 1, pp. 3–9, ISBN: 9781492080640.
- [52] M. Davies, “Concept mapping, mind mapping and argument mapping: What are the differences and do they matter?” *Higher Education*, vol. 62, no. 3, pp. 279–301, 2011, ISSN: 00181560, 1573174X. Accessed: May 7, 2026. [Online]. Available: <http://www.jstor.org/stable/41477852>.
- [53] Å. Wikberg-Nilsson, Å. Ericson, and P. Törlind, *Design: process och metod*. Studentlitteratur AB, 2015, Chapters 1 & 4.
- [54] E. Sanders and P. J. Stappers, “Co-creation and the new landscapes of design,” *CoDesign*, vol. 4, pp. 5–18, Mar. 2008. DOI: 10.1080/15710880701875068.
- [55] J. Patton and P. Economy, *User story mapping: discover the whole story, build the right product.* " O’Reilly Media, Inc.", 2014, ISBN: 9781491904909.
- [56] A. Endmann and D. KeSSner, “User journey mapping a method in user experience design,” *i-com*, vol. 15, no. 1, pp. 105–110, 2016. DOI: doi : 10.1515/icom-2016-0010.
- [57] L.-O. Bligård, *ACD3: Utvecklingsprocessen ur ett människa-maskinperspektiv*. Göteborg: Chalmers University of Technology, 2015. DOI: 10.13140/RG.2.1.1954.4400.
- [58] D. Mcdonagh and I. Storer, “Mood boards as a design catalyst and resource: Researching an under-researched area,” *The Design Journal*, vol. 7, no. 3, pp. 16–31, 2004. DOI: 10.2752/146069204789338424.

- [59] R. Boyatzis, *Transforming Qualitative Information: Thematic Analysis and Code Development*. SAGE, Jan. 1998, ISBN: 0761909605.
- [60] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006. DOI: 10.1191/1478088706qp063oa.
- [61] T. Kravchenko, T. Bogdanova, and T. Shevgunov, “Ranking requirements using moscow methodology in practice,” in *Computer science on-line conference*, Springer, 2022, pp. 188–199. DOI: 10.1007/978-3-031-09073-8_18.
- [62] D. Adams, *The Hitchhikers Guide to the Galaxy*. Pan Books, 1979, ISBN: 9781529046137.

A

Appendix

A.1 Design Criteria

Category	#	Design Criteria
Acessability	1.1	Speech and text (output)
Aesthetic	2.1	Playfullness with words
Aesthetic	2.2	No neutral moderation, No purely technical framing, Humor without trivialization
Aesthetic	2.3	keep (visual) references to HotB
Aesthetic	2.4	not fully realistic = kitsch not enjoyable (abstract ok)
Aesthetic	2.5	These are AI avatars, not historical reenactments, informed about the contemporary world
Aesthetic	2.6	The AI personas are interpretations, not real voices.
Aesthetic	2.7	Audience prompts, interventions are visible, what you experience is curated, performative, and accountable.
Aesthetic	2.8	What really matters is the visual experience, the timing and the excitement generated by the conversational debate
Aesthetic	2.9	The look and feel should resemble a walk-in film, like a montage or collage of events.
Audience emotions	3.1	evoke a sense of "eeriness" regarding AI's role in future decision-making and democracy.
Audience emotions	3.2	Solutions for the future, inspiring
Audience emotions	3.3	The system exposes limits, complicity, and responsibility.
Audience emotions	3.4	encourages curiosity, play, critical thinking, and emotional resonance.
Audience emotions	3.5	makes audience aware of the stakes of technology, ethics, and human choice.
Audience emotions	3.6	Visitors may leave feeling intrigued, unsettled, amused, or challenged

Category	#	Design Criteria
Audience emotions	3.7	engagement should be enjoyable and seductive
Audience emotions	3.8	The atmosphere will be strongly influenced by the quality of the voices or a special moment in the auditory sphere (e.g. a ping, a gong or a single sound as a transition).
Audience interaction	4.1	encourages curiosity, play, critical thinking, and emotional resonance
Audience interaction	4.2	Minimal control on the user from the program to evoke curiosity, play and aha moments
Audience interaction	4.3	the audience can ask anything.
Audience interaction	4.4	Not a knowledge test or a controlled Q&A, but a thinking space.
Audience take-aways	5.1	leave feeling intrigued, unsettled, amused, or challenged
Data/back-end	6.1	They should generate positions from multiple sources
Data/back-end	6.2	Questions recur with altered context, Past answers are quoted back and challenged
Ongoing conversation	7.1	No final answers, no consensus
Ongoing conversation	7.2	Consensus triggers intervention
Ongoing conversation	7.3	Audience questions act as interruptions, not as agenda items.
Ongoing conversation	7.4	The idea is an ongoing, living conversation.
Transparency	8.1	transparency, open to the audience, not a black box.
Transparency	8.2	to recognize that every automated decision is a human choice.
Transparency	8.3	aware of the stakes of technology, ethics, and human choice
Transparency	8.4	making the processes visible, not hidden.
Transparency	8.5	Ideally, it should be akin to reading an absorbing book or watching an engrossing film. Then it becomes transparent.

A.2 Functions and corresponding Solutions to be Combined

#	Function	Solution
7	Controllars	Voice Input, press to speak
9	Controllars	Controller on podium, screen in front
28	Conversation states	Toggle between rastis different conversation "states"
33	Conversation states	All philosophers simultaniously
18	Data recording	text based chat history
23	Input - text	Basically empty input screen, only a prompt and space to write
25	Input - text	Input textruta (output screen) med uppmanande prompt "interrupt, please..."
1	Input/output installation	Seperate input/output screen, text based input, audio based output
13	Map text to character	Text overlapping - grouped with each image of the characters
21	Map text to character	Shows what will happen "flusser is thinking..."
22	Map text to character	column mapping of text and corresponding philosopher
10	Map text to character	Image size/boarder increase when correlating text appears
12	Map text to character	Symbols surrounding portraits + apperaring in text
30	Navigation	Scoll in chat history
2	Output - aesthetics	No text, visual abstraction and audio output
11	Output - aesthetics	Living room visual 2d space
14	Output - aesthetics	aesthetic, symbol based, mappings
20	Output - aesthetics	Traversing space, data points connecting infinitely
26	Output - aesthetics	Each philosopher has a "home", abstraction visualisation
34	Output - aesthetics	4th wall break, "the AIs, seeing the visitor"
3	Output - aesthetics	Big installation with projections on each wall, alien becon
4	Output - aesthetics	Infinite "waterfall" of text from wall to floor
5	Output - aesthetics	Monitor input, dark room + audio output
8	Output - aesthetics	Dual sided screen "backend & frontend"
15	Output - aesthetics	Cords representing flow of information
29	Output - text	Text in the center of UI
32	Output - text	One philosophers text at a time, but dialog between all

#	Function	Solution
35	Output - text	Flowing text, new message on bottom
6	Output - text	Flowing text, new message on top
17	Output - text	Chatroom
24	Output - text	Centered text, and threads connecting to each philosopher
31	System information	Information abt philosophers next to their profiles
16	Thermometer	Visual abstraction of text and intensity of conversation
27	Thermometer	Mapping abstraction of philosophers in a dynamic space
19	Visualizing the user	Camera captures visitor, viz through abstraction together w philosophers

A.3 Initial Concepts for Scenario 2: Wisdome

Visual Concept 1: The first visual concept, as seen in figure A.3.1, incorporated eccentric background visuals, through abstract visual elements combined with the philosophers faces scattered throughout the visualization. This was paired with a contrasting minimalistic text-feed in a code-like aesthetic, where the AI-philosophers output was displayed in an infinite loop of text flowing from bottom to top. The text is white on black background, with four different colors representing the different philosophers.



Figure A.3.1: Visual concept 1: text-based UI, and background visuals in dome-setting

Visual Concept 2: The second visual concept, as seen in figure A.3.2 showcased minimalistic, grid-like background visuals, with more focus on the live-feed screen. Concept 2 featured portraits of the philosopher currently speaking, which swaps when another philosopher gets the word. The live-feed is minimalistic, the text of the current philosopher is centered and moves down as another philosopher is speaking. The text is white on a gray background, with colors representing the different philosophers. The same colors are used for the portraits.



Figure A.3.2: Visual concept 2: text-based UI, and background visuals in a dome-setting

Interaction Concept 1: The first concept for interaction, as shown in A.3.3 featured a mobile application that the users could access on their own mobile devices. This application featured a simple voice recording function, where the user presses and holds button to record their message via voice to the philosophers. If the user lets go of the button, the message will be sent to the live-feed. The app uses simple

white text on gray background, with a big red button for "record", paired with a microphone symbol. This button turns blue when pressed, and the icon changes to a paper airplane, symbolizing the message being sent.

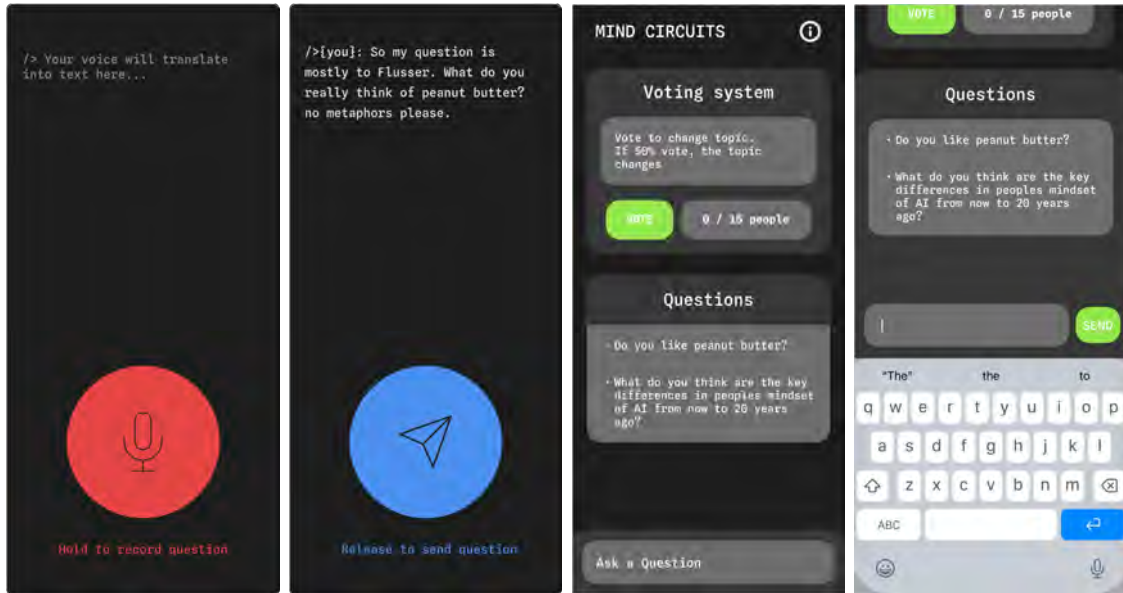


Figure A.3.3: Figure (a), (b) describes interaction concept 1: record and release mobile UI screens. Figure (c), (d) describes interaction concept 2: voting and asking questions mobile UI screens

Interaction Concept 2: The second concept for interaction, as seen in A.3.3 also featured a mobile application which the users could access on their own mobile devices. This concept featured a text-based interaction function, where you are able to type a question to the philosophers. Your typed question then appears in the *Questions* tab, which everyone in the audience can see on their device. A random question from the *Questions* tab would be selected and discussed by the philosophers. The concept also featured a voting system, where if 50% of the audience presses the *Vote* button, the current question discussed would be skipped to another random question. This application features a gray color scale for segmenting different features, with white text and green highlights for important elements.

A.4 Initial Concepts for Scenario 1: Installation

Installation Concept 1: This concept, as seen in A.4.1 featured a separate input-output screen. For the input screen, there is an on-boarding user control screen, with different commands. The user is able to type questions directly to the program using a keyboard. The questions are then displayed on the output screen, which is a large projector or TV screen. The screen features images of all philosophers, with infinitely looping text in the middle of the screen, moving from bottom to top. The input screen has a simple code-like aesthetic, with only white text on gray background. The output screen is more eccentric, featuring white text on gray background, with different colors for the names of the philosophers. This color is also used for the portraits, that each feature a distinct symbol to further differentiate the philosophers.

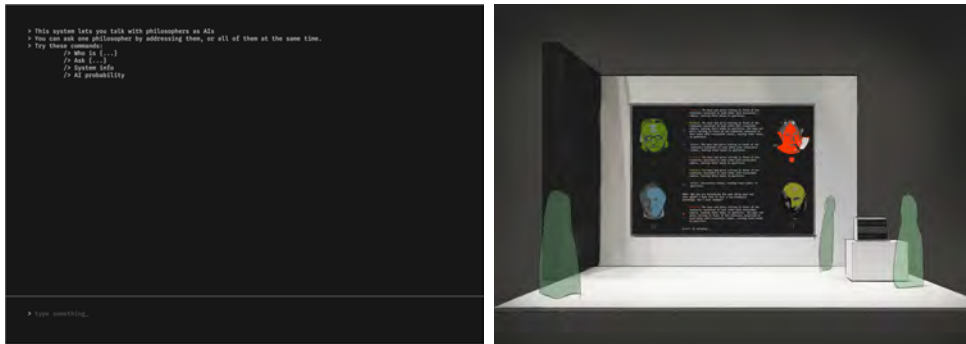


Figure A.4.1: Installation concept 1: Input screen, and exhibition with output screen

installation Concept 2: The second concept, as seen in A.4.2, features a separate input-output screen. For the input screen, the user is able to type their questions directly to the program using a keyboard, which is then sent to the system. The input screen features a code-like aesthetic on a gray background, with different colored text to differentiate user question and philosopher output. The names of the philosophers also have distinct colors. The output-screen is a visual representation of the philosopher currently speaking, presented through abstract shapes and colors. Here, each philosopher also has a distinct shape and color. This visual representation moves when they speak. Based on the intensity of the conversation, the visual representation shows more or less movement.

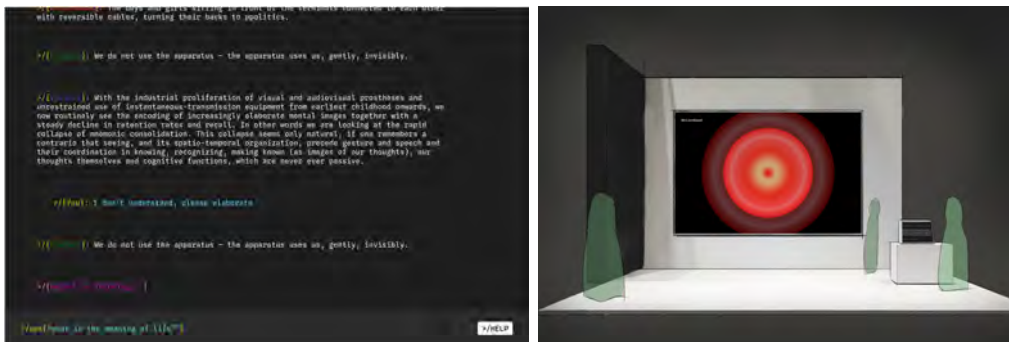


Figure A.4.2: Installation concept 2: Input screen, and exhibition with output screen

A.5 Semi-structured interview guide

This appendix provides the semi-structured interview protocol used during the user evaluation at Wisdome.

Introduction

Describe what the project is:

Mind Circuits is an interactive AI program that brings to life four European media philosophers, whom users can interact and discuss with. With the rapid development and implementation of modern artificial intelligence, users have the opportunity to ask questions or express their concerns/hopes regarding these topics to “AI personalities” whose characters already predicted this modern development some time ago.

Every philosopher will introduce themselves, and then the user has the opportunity to ask a question. The AI philosophers then take a stand on the question, discussing and arguing with each other until the next question is presented.

Explain about Wisdome and what the event will look like.

Start with:

1. *How do you view today's use of AI, specifically LLMs? Do you use it?*
2. *How does it feel to interact with AIs of deceased persons?*

Show Prototype (Semi-structured interview)

Open questions, letting the user guide the conversation → give us insight into focus → potentially change test.

Relevant topics:

- **Portraits:** *Which gives the most immersive feeling of dialogue all at the same time or one-on-one? Since we do not have all four philosophers on screen, is it clear how many are participating in the discussion, and who is speaking?*
- **User question:** *Is it relevant to have it static or should it flow?*
- **Flow of information generally:** *Do you understand who is speaking? Do you understand everyone's individual personas?*
- **“...is thinking”:** *Is it valuable? (It is not included in the UI right now).*

Build up scenario:

“You are listening to this conversation in a spherical room, with a large screen in the middle showing the conversation. Around you and above you, things are displayed, what is it?”

1. *What would you have wanted to see as a complement to this conversation?*

2. *What would have been distracting?*

3. *What would have been immersive?*

Closing

1. *What question would you have wanted to ask?*

2. *How much do you trust this system?*

A.6 Survey Questionnaire Template 1

This appendix provides the first survey template used during the first on-set test of the system with users in Wisdome, Universeum.

Mind Circuits Test Session 13.04.2026

This survey includes questions regarding your experience during this live experiment. In this survey, **Generative AI** refers to systems designed to create original content, such as text, images, or audio, by learning from existing data patterns.

Mind Circuits is an interactive project exploring AI, philosophy and public thinking through the artistic medium and speculative design. Inspired by *Home of the Brain* (1990) by Monika Fleischmann and Wolfgang Strauss, this event re-imagines media art through contemporary AI. At the center of this experiment are four AI-generated personas based on **Joseph Weizenbaum, Vilém Flusser, Paul Virilio, and Peter Weibel**, who debate one another in real time.

The survey consists of **14 questions** and takes about **5 minutes** to complete. Your thoughts and insights are incredibly valuable to our research.

Thank you for your time and your contribution!

Artists: Monika Fleischmann & Wolfgang Strauss · Project Lead: Jasmina Maric, Chalmers · Master's thesis students: Max Hagman, Sally Kalm & Rasti Tengman Mhealden · Industry Partner SVIONOVA AB: Sebastian Norlin, Fredrik Johansson & Sahil Lakhe, Scionova

Ungårsk

Use and perception of AI

1. How often do you use generative AI-tools? *

- Daily
- A few times a week
- A few times a month
- A few times a year
- Never

2. For what do you use generative AI-tools? *

- Information seeking & learning
- Writing & communication
- Personal productivity & planning
- Creative content generation
- Problem solving & technical assistance
- Annat

3. What is your **general position** toward generative AI tools? *

1. Very negative 2 3 4 5 6. Very positive
-

4. Which image set did you prefer? *



Image set 1



Image set 2



Image set 3

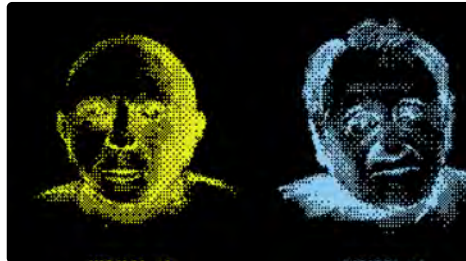


Image set 4

5. Which dialogue style did you prefer? *

- The 1st - Panel
- The 2nd - Debate

6. Describe your experience in **three words**: *

7. Why? Please elaborate (in regards to your answer to question 6)

8. Rate your experience based on these statements *

	1	2	3	4	5	6
How believable was the content provided by the AIs? (1 - not believable at all ; 6 - very believable)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How fluent did you feel the conversation was? (1 - Machine like, not fluent ; 6- Human like, very fluent)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How did the visual elements affect your experience? (1 - Made it significantly worse ; 6 - Made it significantly better)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9. Why? Please elaborate (in regards to your answers to question 8)

10. When you delegate a choice to a machine, do you still feel responsible for the outcome?

- Yes
- No

11. Has this experience changed your perception of AI? Why?

12. Do you feel like you had more to say but no space in the survey? **Write your thoughts here! It helps a lot :)**

13. Would you be open to participating in an interview or future tests of the MIND CIRCUITS project? Write **your email** here and we'll contact you!

14. **GDPR: If you wrote contact information on question 11.**

I consent to the personal data provided in the form above being stored and processed by the master's thesis group for the purpose of conducting studies linked to the master's thesis. I have been informed about the study and consent to participate. I am aware that my participation is entirely voluntary and that I may withdraw my participation at any time without providing a reason. The data will be stored until September 1, 2026, at the latest and will not be shared with third parties. If you have complaints or questions regarding the processing of your personal data, you can contact the thesis group via sallykalm@hotmail.com or maxalfonshagman@gmail.com. My consent below signifies that I choose to participate in the study and agree that the students Sally Kalm and Max Hagman at Chalmers University of Technology may process my personal data in accordance with current data protection legislation (GDPR).

I consent

Det här innehållet har inte skapats och stöds inte av Microsoft. Data du skickar kommer att skickas till formulärets ägare.

 Microsoft Forms

A.7 Survey Questionnaire Template 2

This appendix provides the second survey template used during the second on-set test of the system with users in Wisdome, Universeum.

Mind Circuits Test Session 27.04.2026

This survey includes questions regarding your experience during this live experiment. In this survey, **Generative AI** refers to systems designed to create original content, such as text, images, or audio, by learning from existing data patterns.

Mind Circuits is an interactive project exploring AI, philosophy and public thinking through the artistic medium and speculative design. Inspired by *Home of the Brain* (1990) by Monika Fleischmann and Wolfgang Strauss, this event re-imagines media art through contemporary AI. At the center of this experiment are four AI-generated personas based on **Joseph Weizenbaum, Vilém Flusser, Paul Virilio, and Peter Weibel**, who debate one another in real time.

The survey consists of **9 questions** and takes about **5 minutes** to complete. Your thoughts and insights are incredibly valuable to our research.

Thank you for your time and your contribution!

Artists: Monika Fleischmann & Wolfgang Strauss · Project Lead: Jasmina Maric, Chalmers · Master's thesis students: Max Hagman, Sally Kalm & Rasti Tengman Mhealden · Industry Partner SVIONOVA AB: Sebastian Norlin, Fredrik Johansson & Sahil Lakhe, Scionova

Ungårsk

1. How often do you use generative AI-tools? *

- Daily
- A few times a week
- A few times a month
- A few times a year
- Never

2. For what do you use generative AI-tools? *

- Information seeking & learning
- Writing & communication
- Personal productivity & planning
- Creative content generation
- Problem solving & technical assistance
- Annat

3. What is your **general position** toward generative AI tools? *

1. Very
negative

2

3

4

5

6. Very positive

4. Describe your experience in **three words**: *

5. Why did you chose these words? Please elaborate (in regards to your answer to question 4)

6. Did you feel the content provided by the AIs was **believable**? Why so?

7. How did the **visual elements** affect your experience?

8. Did you feel the conversation was **fluent**? (Machine-like or human-like) Describe why!

...

9. How would you describe the experience of interacting with the AI-thinkers?

10. Do you feel like you had more to say but no space in the survey? **Write your thoughts here!
It helps a lot :)**

A.8 Survey Questionnaire Template 3

This appendix provides the third survey template used during the final presentation of MIND CIRCUITS in Wisdome, Universeum.

Mind Circuits Test Session 06.05.2026

This survey includes questions regarding your experience during this live experiment. In this survey, **Generative AI** refers to systems designed to create original content, such as text, images, or audio, by learning from existing data patterns.

Mind Circuits is an interactive project exploring AI, philosophy and public thinking through the artistic medium and speculative design. Inspired by *Home of the Brain* (1990) by Monika Fleischmann and Wolfgang Strauss, this event re-imagines media art through contemporary AI. At the center of this experiment are four AI-generated personas based on **Joseph Weizenbaum, Vilém Flusser, Paul Virilio, and Peter Weibel**, who debate one another in real time.

The survey consists of **12 questions** and takes about **7 minutes** to complete. Your thoughts and insights are incredibly valuable to our research.

Thank you for your time and your contribution!

Artists: Monika Fleischmann & Wolfgang Strauss · Project Lead: Jasmina Maric, Chalmers · Master's thesis students: Max Hagman, Sally Kalm & Rasti Tengman Mhealden · Industry Partner SVIONOVA AB: Sebastian Norlin, Fredrik Johansson & Sahil Lakhe, Scionova · Music by Jakob Dahl

Ungatoriski

1. How often do you use generative AI-tools? *

- Daily
- A few times a week
- A few times a month
- A few times a year
- Never

2. For what do you use generative AI-tools? *

- Information seeking & learning
- Writing & communication
- Personal productivity & planning
- Creative content generation
- Problem solving & technical assistance
- Annat

3. What is your **general position** toward generative AI tools? *

1. Very negative	2	3	4	5	6. Very positive
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

4. Describe your experience in **three words**: *

5. Why? Please elaborate (in regards to your answer to question 4)

6. Rate your experience of the philosophers based on these statements *

	1	2	3	4	5	6
How believable was the content provided by the AIs? (1 - not believable at all ; 6 - very believable)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How fluent did you feel the conversation was? (1 - Machine like, not fluent ; 6- Human like, very fluent)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How consistent was the AI characters' personalities throughout the interaction? (1 - felt like disconnected AI responses , 6 - felt like a continuous presence)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7. Why? Please elaborate (in regards to your answers to question 6)

8. Rate your experience of the interaction based on these statements *

	1	2	3	4	5	6
How did the visual elements affect your experience? (1 - Made it significantly worse ; 6 - Made it significantly better)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How understandable was the interaction with the system? (1- hard to follow , 6 - easy to follow)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How did you perceive the arrangements of visual elements ; names, subtitles, questions, portraits? (1 - distracting, unclear , 6 - purposeful, logical)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9. Why? Please elaborate (in regards to your answers to question 8)

10. When you delegate a choice to a machine, do you still feel responsible for the outcome?

- Yes
- No

⋮

11. Has this experience changed your perception of AI? Why?

12. Do you feel like you had more to say but no space in the survey? **Write your thoughts here! It helps a lot :)**

A.9 State-changing Functions & updated Story Mapping

Table A.1: New state changing functions & their relevance in different use-states

Defining state-changing functions			
	1. ask new question	2. debate	3. follow-up
F	<p>1.1 Q streamed Word-by-Word: <i>If not, Q shows up after finished speaking; Q must fit on screen</i></p> <p>1.2 If a previous AI has spoken: text should not stay on screen</p> <p>1.3 AIs should stop talking / moving</p> <p>1.4 Separate front-end for text input (and controls)</p>	<p>2.1 Separate front-end: <i>Round-up; Pause; Deactivate Gif:s; Hard Reset; Clear screen; Un-mute / mute microphone input</i></p> <p>2.2 X 2.2 Show who is speaking with text & GIF: GIFset based on distance</p>	<p>3.1 System should pause / stop text + voice when new Q is asked</p> <p>3.2 Show new Q in place of last</p> <p>3.3 If question is very wrong → manually pause through separate front-end</p>
B	<p>1.5 Unmuted system should: <i>Pick up audio as input; Send input → backend</i></p> <p>1.6 Word-by-Word: <i>Set delay in back-end</i></p> <p>1.7 Reset Question: <i>Both from UI & back-end</i></p>	<p>2.3 Debate until finished OR interrupt</p> <p>2.4 Round-up: Conversation rounds up → Weibel has last words</p> <p>2.5 Pause: <i>Stop after current AI finished speaking</i></p> <p>2.6 Hard reset: <i>Stops AI talking immediately; Flushes convo; Keeps history, keeps text on screen</i></p> <p>2.7 AIs should debate the recent Q: <i>Conversation / debate style</i></p>	<p>3.4 At end of debate, soft Reset should: Flush convo; Keep memory, but make sure we leave topic</p>

Additionally, the use-state *0. audience enter* included a need for the front-end to show a black screen until boot-up sequence is toggles, and the back-end needed to make the system mutable through keyboard-shortcuts and menu controls. The *4. end* included a need for a end-credit screen with credited names and a survey QR-code.