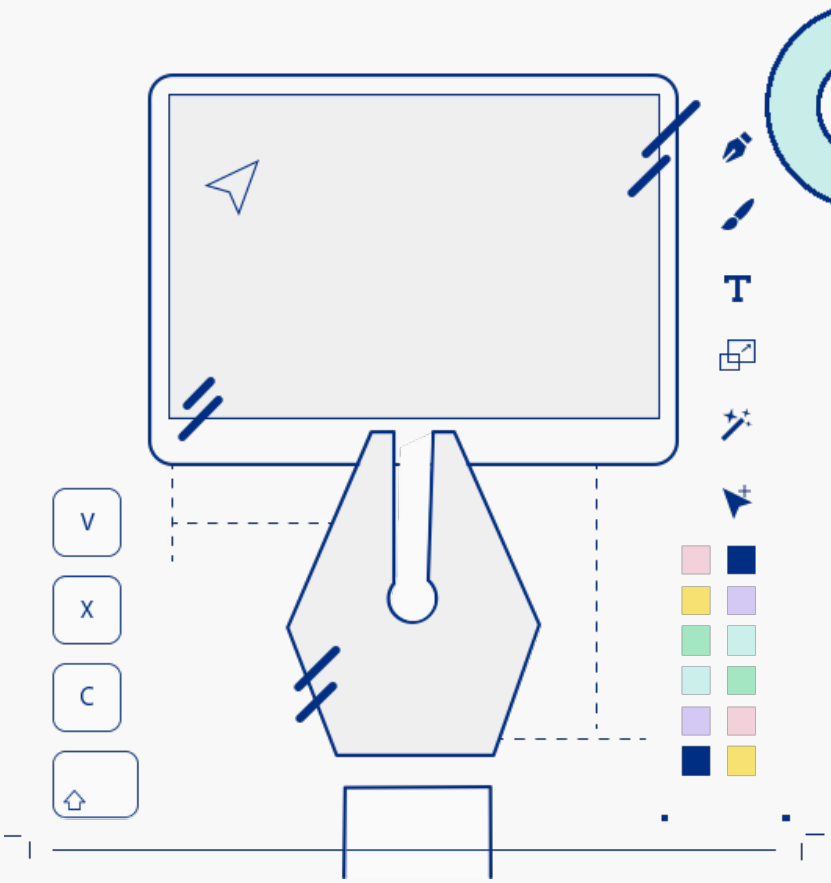
Develop an interface for model-informed prototyping of HAI interactions



Graduation Master Thesis

Jiayi Zhou
Design for Interaction
Student #5568935

Supervisor team:
Dave Murray–Rust IDE/HCD Chair
Mahan Mehrvarz IDE/HCD Mentor
Konstantinos Tsiakas IDE/HCD 2nd Mentor

Summary

As AI technology continues to advance, there's a growing need to integrate it into UX design. However, AI's unique characteristics does not seamlessly align with current design tools, and mastering the technical aspects for designers is a significant challenge. The project goal is to develop a tool based on a developed semi-formal representation for Human-AI (HAI) interactions, which uses a set of communicative acts[1] to specify the communicated information between users and AI models as exchanges of messages. It followed design considerations which referred to those for the Model-Informed Prototyping[2](MIP). See *Figure c*.

The project followed an iterative prototyping method [10, 11] (*Figure a*) across four phases to get insights or assess ideas for the final design output:

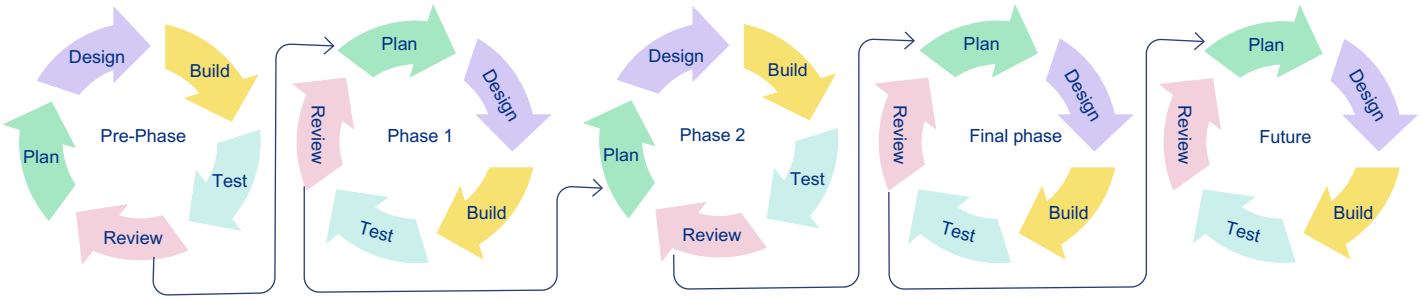


Figure a. Iterative prototyping method process.

– **Pre-Phase**: Tested the understanding of communicative acts with 2 design students using paper materials and a use case ("CV-Screening"). Insights from this phase guided future design considerations, and suggestions for improving the use case and the data structure of the final design output were noted. See *Figure b & c* for the overview of insights.

Suggestions for Design #1	Insights on structures – The linear structure is preferred
Suggestions for Design #2	Insights on Design considerations
Suggestions for Design #3	Communicative acts make sense for participants

Figure b. Overview of insights for the Pre-Phase.

1 Communicative acts help humans communicate with AIs by exchanging specific types of information.
2 Model-Informed Prototyping (MIP), a workflow that combines model exploration and interface design tasks [14].

Design Considerations for model-informed prototyping [13]	Design Considerations for the project	Possible functions
Prototyping tools should allow designers to invoke ML models by specifying input data directly.	1. Designers can create, modify, delete and move messages to form interactions. 2. They can modify the instances or data information about inputs/outputs/instances, etc. of the AI models according to the needs of the end–users for evaluating the predictions.	Create & modify messages & message sequences; choose the best AI models based on the end–users' needs; connect instances in the use case with messages; visualize the input data during the HAI…
Prototyping tools should allow designers to incorporate AI outputs into interface design.	Designers shall be able to visualize in the interface the instances of each Message that are related to the exchange of information or the HAI–related UI elements.	Visualize the output & feedback/XAI; connect instances in the use case with messages; choose the best AI models based on the end–users' needs…
Prototyping tools should allow designers to shape model APIs according to end–user needs.	Designers can define the inputs/outputs/Feedback–XAI of the AI model and the presentation of these based on the needs of the end user in the use case. However, they should not involve too much coding and focus more on how the designer builds the HAI.	Choose the best AI models based on the end–users' needs; help designers understand functions of each AI model easily…
Prototyping tools should allow designers to evaluate design choices across diverse users and contexts.	1. Designers can create personas that will perform HAI based on use cases, whether they are humans or AI models. 2. The design outputs should be broadly applicable to different design challenges.	Create & modify messages & message sequences; connect the end–users' needs with the AI models and messages…
Prototyping tools should allow designers to incorporate model–related data rapidly and iteratively.	Designers should have the flexibility to adapt the content created in the design output to the needs and feedback of the end user.	Create & modify messages & message sequences; visualize the data or instances conveying during the message sequence…

Figure c. The project's design considerations, one of insights in Pre–phase.

– **Phase 1**: Created a low–fidality digital prototype in Figma[1], using the improved "CV–Screening" as a case study. 6 participants explored effective ways to present communicative acts and strategies for representing Message sequences in Human–AI Interaction. This phase provided precise design goals, generated design ideas, and refined the use case. See *Figure d* for the overview of insights.

Suggestions for Design #1	Emphasize and clearly demonstrate the connection between Instance and Terms.
Suggestions for Design #2	Fewer interfaces would be better.
Suggestions for Design #3	Be careful to clearly distinguish and present content between different levels in the prototype.
Suggestions for Test #1	Providing a more reasonable test context and supporting materials in the subsequent tests.

Figure d. Overview of insights for the Phase 1.

1 Wikipedia. (2023, September 27). Figma. https://en.wikipedia.org/wiki/Figma

– **Phase 2**: Built a high–fidelity prototype based on user journey map. Test 2 assessed if the design prototype met design goals and design considerations. Two participants provided insights for further improvements on both design concepts and the test materials. See *Figure e* for the overview of insights.

General insight in the design & test plan	The digital prototype and the test plan worked in general
Suggestions for Test #1	The analysis of two pilot tests highlights the necessity for a clearer introduction
Suggestions for Test #2	Display all the parts that can be interacted with where they can be seen most easily, while ensuring that interactions do not interfere with task understanding
Suggestions for Test #3	Reorganize the questions in each task so that they encourage participants to interact with the prototype
Suggestions for Design #1	Change some descriptions in the prototype
Suggestions for Design #2	Keep interactions for the same purpose the same and logical
Suggestions for Design #3	Suggestions on UI components for the prototype

Figure e.. Overview of insights for the Phase 2.

– **Final Phase**: Used the refined digital prototype for the "CV–Screening" use case in the last test, offering crucial insights for future project development. See *Figure f* for the overview of insights.

Good aspects of the design concept #1	Familiarity helps designers more easily understand the concepts in the project
Good aspects of the design concept #2	Simplicity helps designers focus more
Suggestions for what could be improved #1	Organization of content in Messages
Suggestions for what could be improved #2	Explore diverse structures beyond the linear format
Suggestions for what could be improved #3	Touchpoint's content has too many details in UI components

Figure f. Overview of insights for the final Phase.

The final output of the project is a partial prototype of a digital tool designed to facilitate the early stages of human–AI interaction design (*Figure g*). Grounded in the principles of communicative acts and human–centered design, this tool assists designers during the Ideation stage of Design process. It achieves this by visualizing the roles, data, and information involved in the process of information exchange during Human–AI Interactions. The goal is to enhance efficiency and ease in designing these interactions.

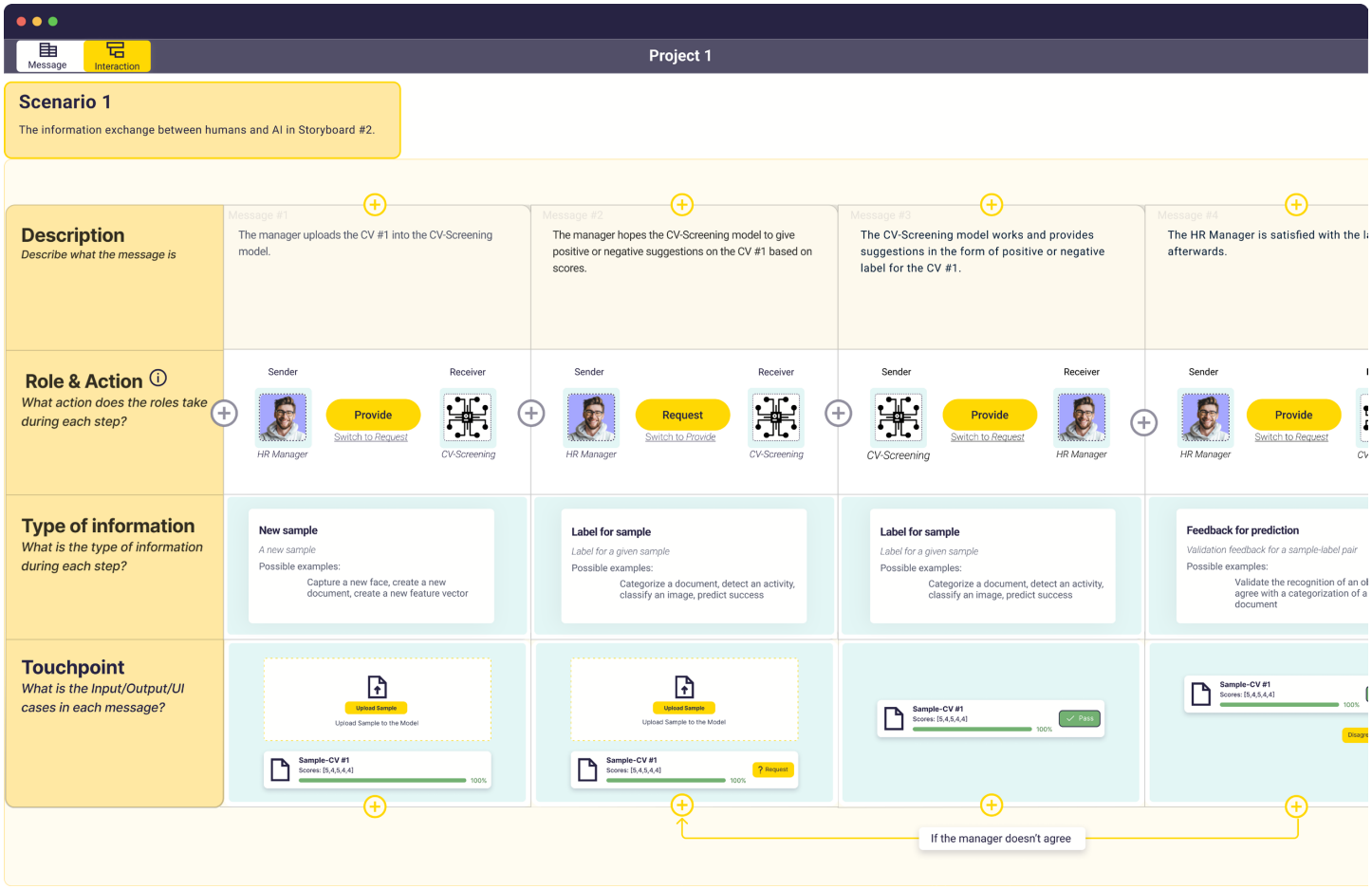


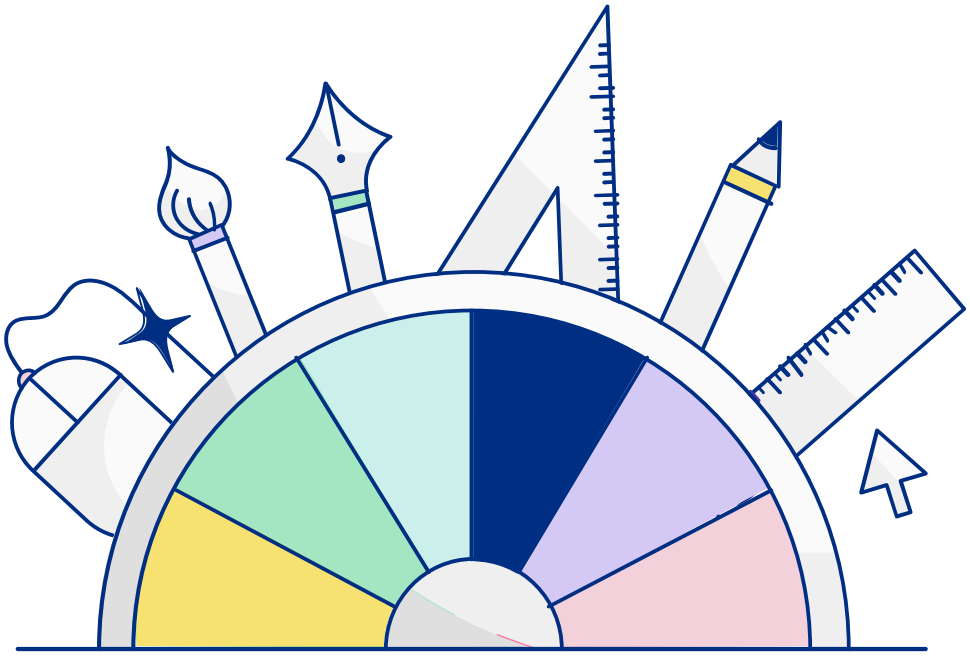
Figure g. Example of the final prototype in Figma.

Table of content

1.

Introduction

The first part is all about giving essential details regarding the project, covering aspects like why designing such a tool is necessary, who would find it useful, and what it should be like. Find the original project brief in Appendix A.



1.1 Why designing HAI is challenging?

1.1.1 Background

Advancements in technology steadily brings more possibilitie for innovative, and designers play an important role in using new technologies to make life better. Their ability to learn and deconstruct cutting–edge technologies allows them to apply these insights across various fields, contributing to progress on a broader scale.

In the area of technological evolution, Artificial Intelligence (AI) has developed for almost a century, holding a large accumulation of technologies and extensive data resources [1]. This transformative field has now captured the keen interest of designers [2, 3, 4, 5, 6]. This surge in scholarly exploration is indicative of the significant role AI may play as a new material in augmenting user experiences [5]. However, within the domain of user interface and user experience design, practitioners are facing challenges as they try to find a way to effectively integrate AI as a foundational element in UX design [5, 7, 8].

While exploring the AI models, designers face many challenges. For example, technical complexities and uncertainties [8]. Designers, especially those without specialized AI training, face difficulties in creating effective interface prototypes that integrate AI into UX design. Notably, current interaction prototyping tools which are popular, like Figma, are primarily focused on crafting traditional interaction interfaces, and may not fully address the unique characteristics of AI. Some other tools, like Voiceflow, cater to specific AI domains. The combination of technical complexity and uncertainty highlights the potential limitations of traditional prototyping methods in capturing the full behavior and capabilities of AI systems in the context of user experiences [8].

The primary goal of this project is to develop a prototyping tool for HAI interactions based on the semi–formal representation of HAI interaction as sequence of messages. The messages can describe the communication of specific types of information between users and models [9]. See *Figure 1.1–1.5*.

Figure 1.1 provides an overview of the communicative act. It shows how specific types of information exchanged between humans and AI models. It is the process by which humans and AIs engage in communication from a systems perspective.

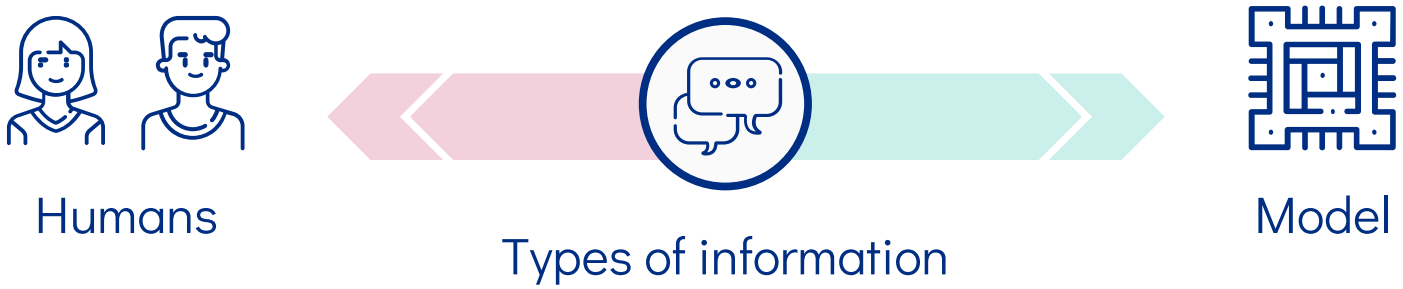


Figure 1.1. Overview of the communicative act.

Humans and AI models interact with each other through messages. More specifically, a sender communicates with a receiver by providing or requesting a specific type of information. Either humans or AI models can be Senders or Receivers. As shown in *Figure 1.2*.

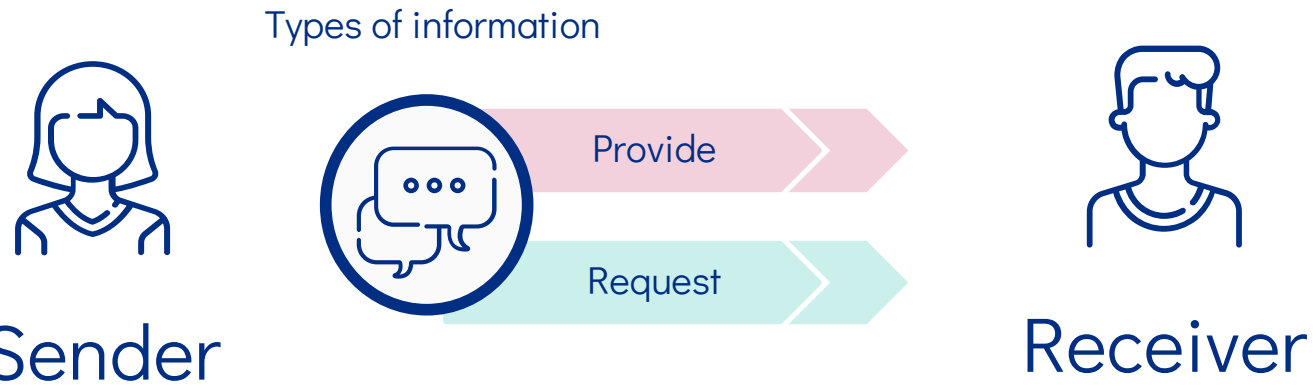


Figure 1.2. One Message: the sender communicates with the receiver by either providing or requesting a specific type of information.

And a sequence of messages can describe an interaction scenario between specific Human users and AI models, which is know as Human–AI Interaction. It is shown in *Figure 1.3*. More discussion about Human–AI Interaction is in Chapter 2.

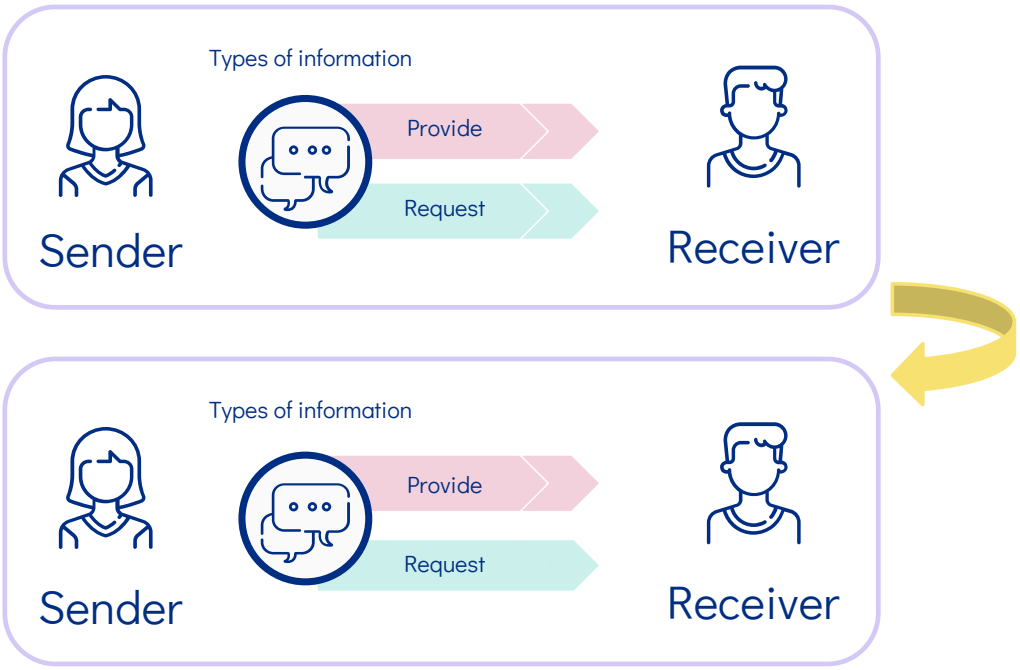


Figure 1.3. A sequence of messages can describe a human–AI interaction.

We have defined a set of 12 types of information, based on the model data types (input, output and feedback/XAI). *Figure 1.4* shows the overview of 12 types of information. They are the model–based information exchanged in communication between humans and AI models.Overall, they cover the following three areas:

– model input: data types used from the model as an input, e.g., a CV document
– model output: data types used from the model as an output, e.g., a CV label
– explanations and user feedback: data types used from the model as additional feedback/XAI, e.g., validation feedback, explanations

For more details, see Appendix B.

Figure 1.4. Overview of 12 types of information.

And the human and the AI models respectively output or receive these 12 types of information through their own understandable ways. See *Figure 1.5*. For AI models, they give out Output represented by Prediction, Classification, etc. and Feedback represented by Explanation, while they receive Input from external sources (which may be humans or other AI models) such as Data or Symbols. But these are mainly composed of machine language and are not readable for humans. Therefore, for humans, they mainly exchange information with the AI model through UI elements in the interface. Specifically, the humans will get the content of the AIs output from the information presented by the UI, and then feedback the information through the UI elements.

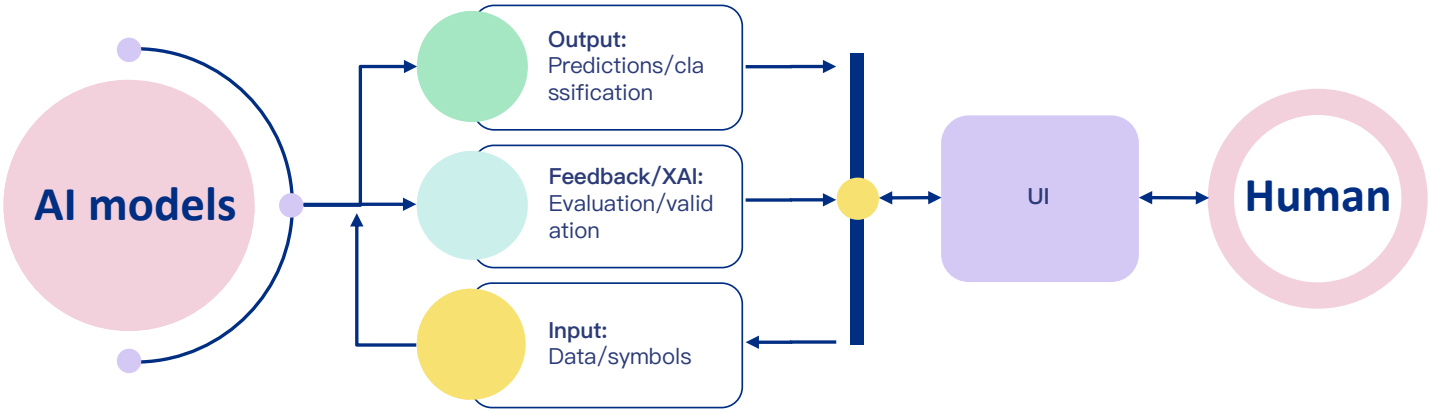


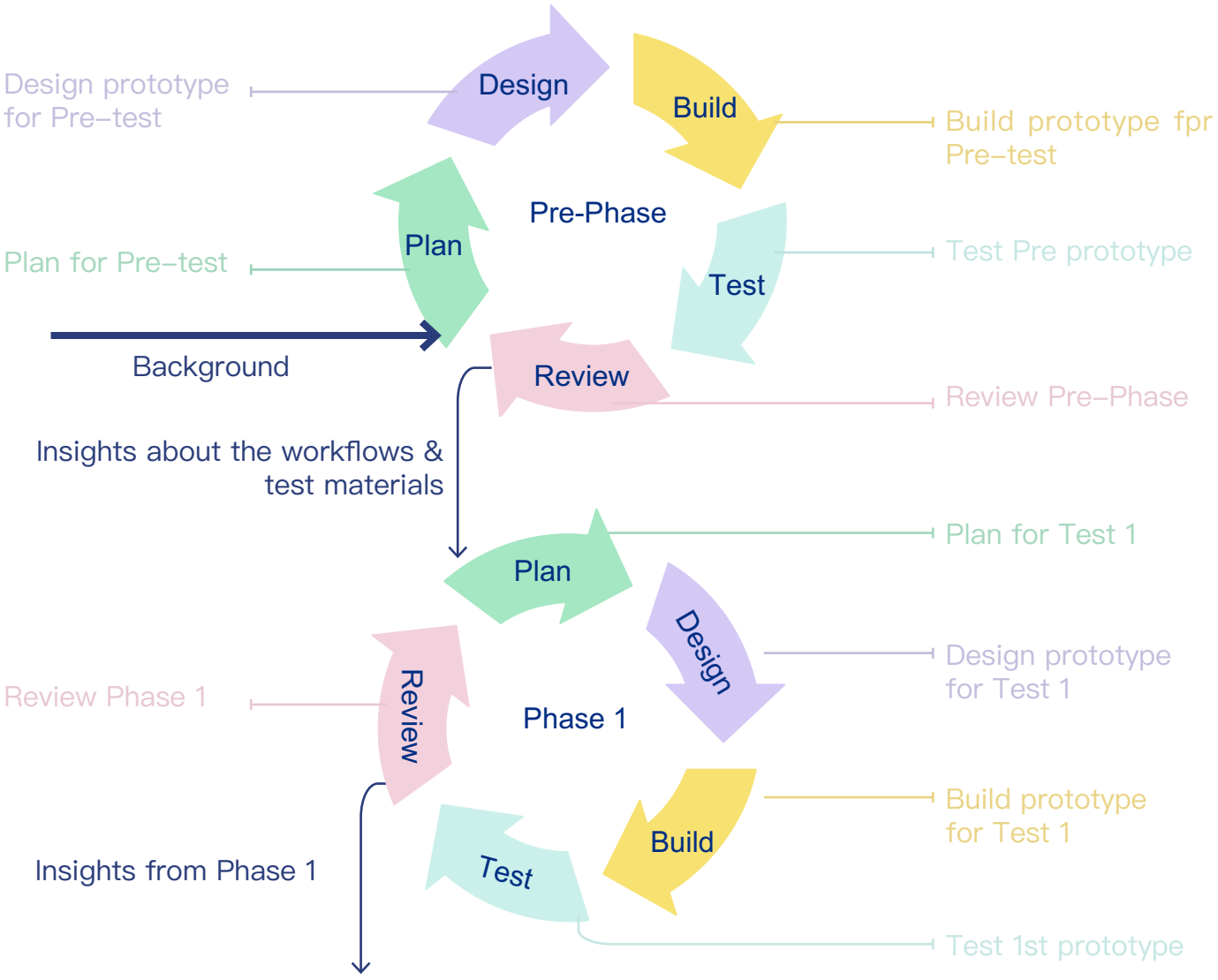
Figure 1.5. Details on information a person or AI requests or provides.

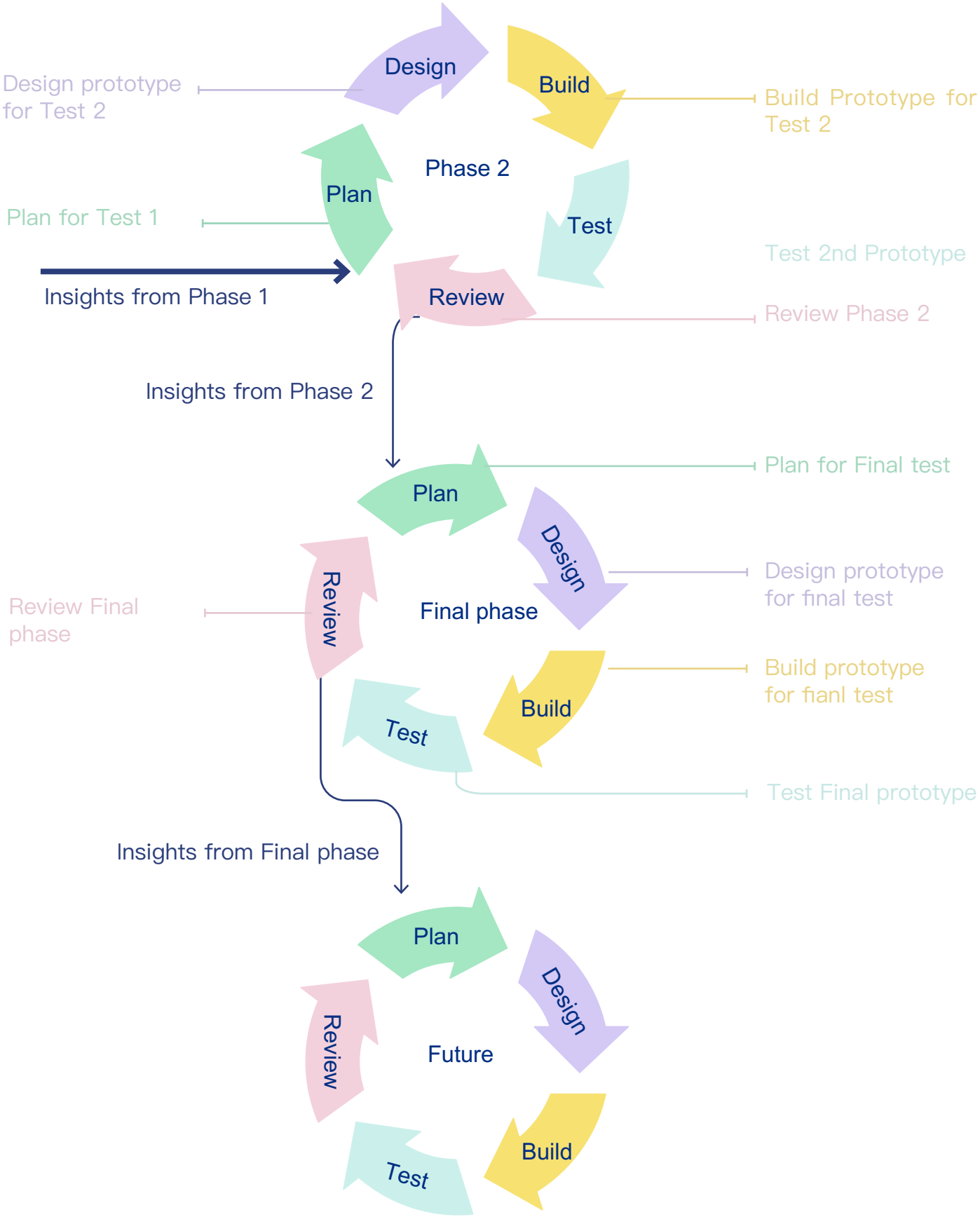
In conclusion, this project aims to contribute to the integration of design within the AI technology landscape, creating a stronge connection between design and engineering realms.

1.1.2 Project method

The primary goal of this project is to develop a prototyping tool for HAI interactions based on the semi-formal representation of HAI interaction as sequence of messages (see *Figure 1.1 – 1.5*). An iterative prototyping methodology [10, 11] was used throughout the design and research process, with each phase consisting of the following steps: analyze, plan, design, build, test, review, and release. *Figure 2* illustrates how the iterative prototyping method was used as a guide to carry out the activities in this project. For explanations on each phase, please see the page after *Figure 2*.

Figure 2. Iterative prototyping approach and activities in the project .




Design prototype for Test 2
Design
Build
Build Prototype for Test 2
Phase 2
Plan
Plan for Test 1
Test
Test 2nd Prototype
Insights from Phase 1
Review
Review Phase 2
Insights from Phase 2
Plan for Final test
Plan
Design
Review Final phase
Review
Final phase
Design prototype for final test
Build prototype for fianl test
Build
Test
Test Final prototype
Insights from Final phase
Plan
Design
Review
Future
Build
Test

Pre–Phase: This stage aimed to fast gather essential insights from design students regarding prototyping tools. 2 participants joined in activities such as educating and creating HAI interactions through paper prototypes. Those with past experience in using HAI as a UX material contributed important insights to inform subsequent prototyping development. Ideally, the phase concluded with design considerations or assessments for the project. More details in Chapter 3.

Phase 1: This phase built on the insights gained in the previous phase and used Figma as the prototype building tool to brainstorm ideas to get an draft version of the digital prototype. The goal in this phase was to gain insights like the digital prototyping workflow and specific design goals. See Chapter 3 for details.

Phase 2: This phase was based on the insights gained in the previous two phases, which resulted in the output of three ideas, which were compared and then selected to be inspired by the User journey map, and interactive models were created for testing using Figma. The test in this phase was done not only to gather insights on how to enhance the existing concepts, but also to evaluate the test materials used for the final testing. See Chapter 5 for more details.

Final Phase: Finally, based on the insights gained from the previous steps, the test materials as well as the digital prototypes for testing were re–improved. The final test of the methodology and functional modules for the development of screen–based tools for the original HAI framework developed by the client. See Chapter 6 for more details.

1.2 Who would find it useful

1.2.1 Target groups

The project brief clearly defines the primary target audience: designers. It means a wide group, including senior designers and design students currently under training.

While the expect is to create a final design output that caters meaningfully to a diverse range of designers, the reality of limited time and the early stage of the project, it is necessary to focus on design students for testing at various stages. This pragmatic approach acknowledges the evolving characteristic of the project and the potential for future enhancements.

Furthermore, recognizing that design is collaborative, involving coordination with project managers, developers, and other stakeholders, design students remain the primary focus for testing at all stages. Although the tool may not be intended specifically for their use, their presence is considered during the developmental phases, anticipating potential adjustments when the tool is ready for broader application.

1.3 Possible design outputs

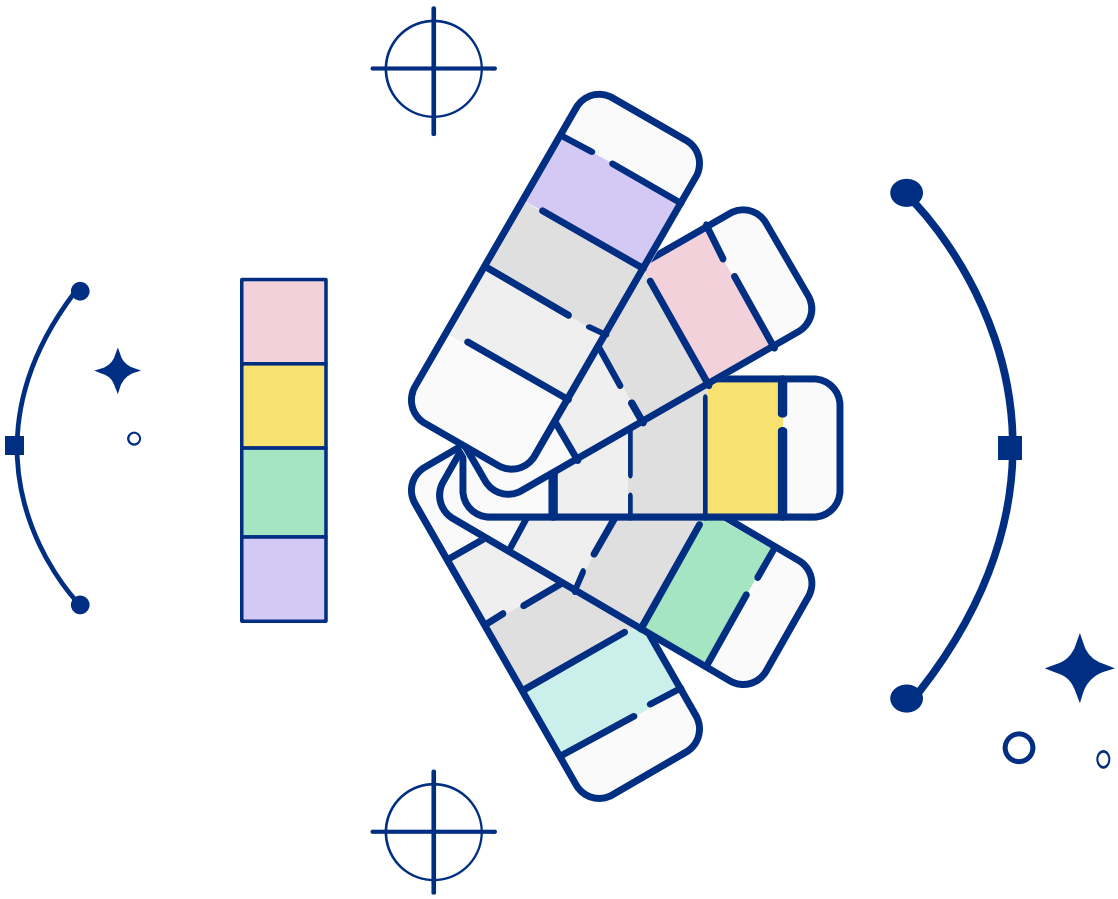
1.3.1 Possible design outcome

The project started with an open-minded approach towards the final outcome, and through numerous meetings and discussions, a concrete decision was made to develop a digital prototyping tool. The primary goal of this project is to develop a prototyping tool for HAI interactions based on the semi-formal representation of HAI interaction as sequence of messages.

1.3.2 Project scopes

Although it was originally planned that the second and third phases would use the programming languages html/css/javascript to create interactive models, due to scheduling and the limited personal programming skills of the researchers, only static interfaces were implemented. To see examples of some of the interfaces, check out: Interface

Based on this, the final prototype was completed by Figma. The link for the prototype: Final prototype



2. Related work

The second section holds related research that has significantly influenced the problem definition of this project, design considerations, and the research methodology. The integration of Artificial Intelligence (AI) as a design material within the area of User Experience (UX) design is a well-explored subject, widely explored by various researchers. These studies have not only generated valuable design insights but have also documented and highlighted the prevalent design challenges [10, 12, 13, 14].

2.1 AI Challenges in Design

Before going further into the research, it is important to realize what AI is and why designing for it has a lot of big challenges. This understanding will lay the groundwork for exploring the challenges of AI for design innovation.

In the discussion about Human–AI Interaction, the definition of AI does not receive a lot of attention, often based on terms associated with machine learning systems to provide a vague explanation without a definitive conclusion [8]. Despite this ambiguity, these discussions offer insights into defining AI, especially within the context of Human–AI Interaction.

One common explanation of AI says it is the set of techniques and methods, including machine learning, used in computer vision, natural language processing, and others [8,15]. This explanation fuels debates within computer–related research fields about how to precisely define AI [16,17]. For example, one short definition of AI in computer–related research fields is“Giving computers the ability to learn without being explicitly programmed”[15].The understanding of AI by designers becomes particularly importantl in this context, given that the project is specifically aimed at designers.

Traditionally, designers play an important role in advancing society by exploring the capabilities and limitations of existing technologies to create valuable designs. Interestingly, in the area of Human–AI Interaction, while researchers in computer–related fields focus on realizing new technological revolutions, designers tend to shift their attention away from the capabilities and technological limitations of AI [8]. Designers, a little similar to developers, are increasingly engaged in designing functional modules for different contexts by constructing the underlying logic of AI modules. Recent research,using strategies like the Wizard of Oz approach, investigates whether designers can grasp the limitations and frontiers of AI as a UX material through well–constructed testing programs [8, 18].

2.2 Human–AI Interaction (HAI)

The definition of HAI is a little similar to that of Human–Computer Interaction [19]. Some studies [8, 20] do not distinguish between HCI and HAI when discussing relevant aspects; these studies consider HAI as a specific type of human–computer interaction and then explore the design challenges facing HAI. Other studies have argued [21, 46] that HAI in the context of AI applications involves active collaboration between human domain experts and AI methods to improve the transparency, accuracy, and credibility of AI results.

According to Chatgpt's answer to the question "What is Human–AI Interaction" [22], Human–AI Interaction (HAI) refers to the interaction and control of humans with AI technology, as well as the ways in which AI systems are used and how they are used. For humans, it includes the way in which the AI is accessed and controlled; for AI models, it includes how humans' input is accessed and results are presented, etc. Specifically, it helps to realize the communication and cooperation between humans and AI through, for example, the development and design of interaction interfaces.

In this project, we follow the semi–formal representation for HAI interactions, where interactions can be described as a sequence of messages between human users and AI models. See details in Chapter 1 and *Figure 1.3*.

One of the project goals is to assess if designers can understand the underlying concepts of the semi–formal representation, including messages, actors, provide/request, and types of information.

2.3 The UX challenges for HAI

By summarizing the problems from related research, it concludes that two important and unique user experience challenges in building HAI are from: 1) the inner uncertainty of AI itself [8, 23], and 2) the complexity of AI's own technologies and concepts, which include, among other things, the dynamics of AI systems [8, 24]. The following shows the detailed design challenges contained in each of these two areas.

The inherent uncertainty of AI itself

– **Understanding AI Capabilities** [5, 23]: Designers often find it hard to understand the capabilities and limitations of AI. This lack of understanding becomes a obstacles in the stages of brainstorming and ideation.

– **Ideating New AI-related Interactions** [5, 25]: Even when designers have knowledge of how AI functions, assuming numerous new and achievable AI interactions for a specific UX problem proves to be difficult. The adaptability and fluidity of AI-powered interactions add complexity to the ideation process.

The complexity of AI's own technologies and concepts

– **Iterative Prototyping and Testing** [26, 27, 28]: Traditional Human-Computer Interaction (HCI) practices, centered around rapid and iterative prototyping, face limitations when applied to AI. The unpredictable nature of AI systems makes it hard to anticipate and fully assess their consequences, which stops effective iterative test.

– **Designing Thoughtful Interactions** [5]: Designers face difficulties in setting appropriate user needs for AI's sometimes unpredictable outputs. Additionally, ethical considerations and concerns about societal consequences associated with AI–powered interactions pose challenges.

– **Collaboration with AI Engineers** [28, 29]: The limited knowledge about AI capabilities and challenges in establishing effective collaboration between designers and AI engineers block the seamless integration of AI into the design process.

– **Adapting Human–Centered Design for AI** [30,31]: Some argue that traditional Human–centered design approaches need to evolve to effectively incorporate AI. The dynamic and complex nature of AI systems may require a fundamental shift in design methodologies.

– **Lack of consensus on underlying causes** [8]: Altoughh there are a large nember of rearches on exploring the challenges for HAI [5,23, 24,29], there is still a lack of unity on the root causes of the various challenges facing AI–centered UX design based on the difficulty to summarize.

2.4 Design thinking

Given that the primary target audience for this project is designers, it is critical to understand how they are thinking about during designing.

Design thinking [32, 33], is commonly working as an analytical and creative process. This method engages individuals in exploring possibilities, fostering researches, and creating prototypes. It operates through an iterative cycle that includes collecting feedback and redesigning based on insights gained from the process. This non–linear, iterative approach, consisting of five phases—Empathize, Define, Ideate, Prototype, and Test—is useed by teams to understand users, challenge assumptions, redefine problems, and develop innovative solutions for prototyping and testing. It is particularly effective for addressing problems that are defined poorly or unknown. See *Figure 3*.

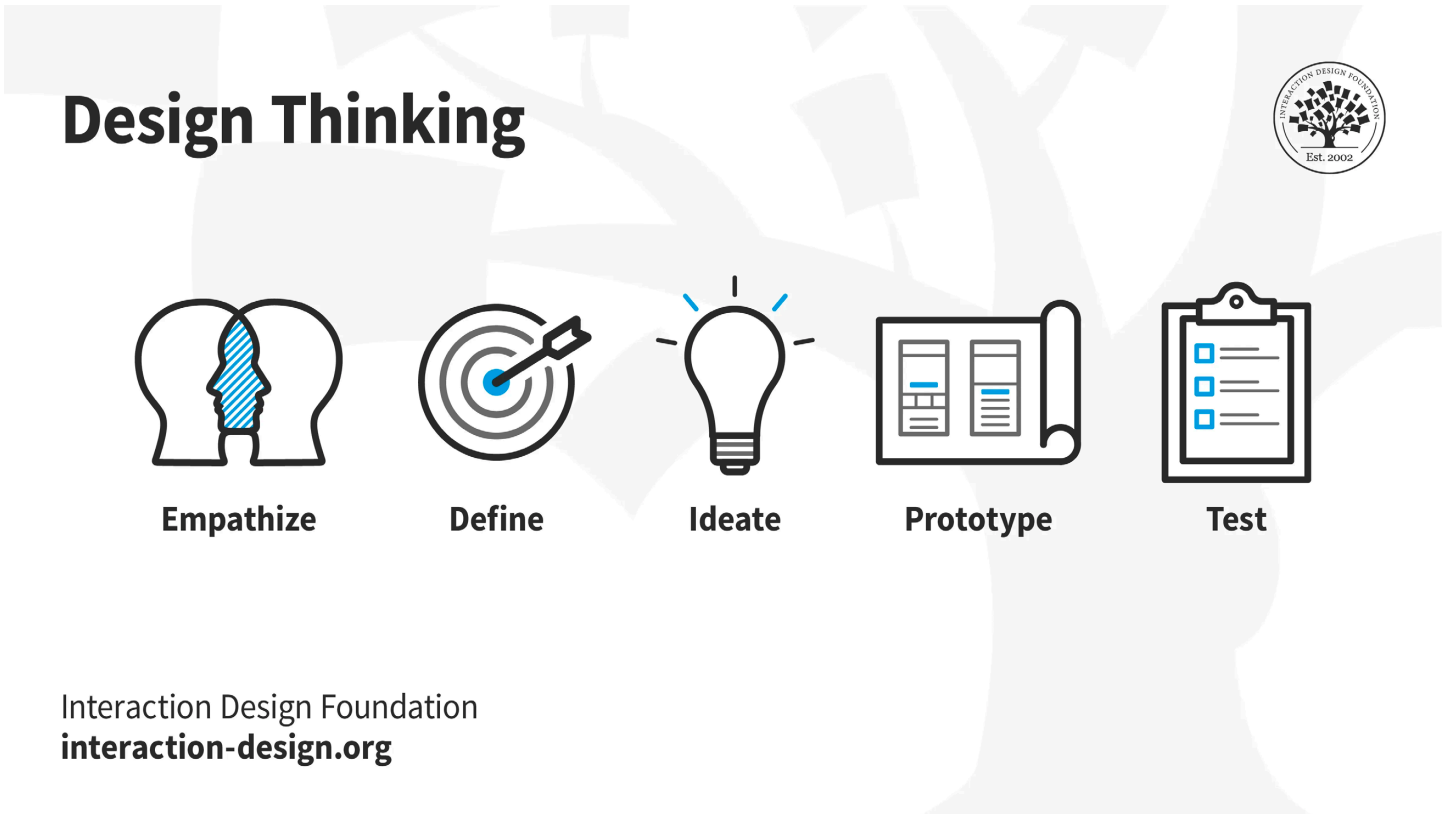


Figure 3. Visulization the Design thinking. From © Interaction Design Foundation.

Design teams frequently use design thinking to address design challenges [33]. This thinking enables them to reframe problems with a human–centered approach, finding aspects that are the most important for users. With this perspective, design teams can adeptly navigate the processes of UX research, prototyping, usability testing, and more, effectively discovering innovative ways to meet users' need.

It is important to note that Design thinking here is a discussion of the general process of designers thinking about design challenges and does not represent the research activities that held in this project.

2.5 Related work

The current products about AI & UX on the market in the field of Human–AI Interaction are valuable for this project. These products help designers gain a deeper understanding of the challenges faced when using AI as UX material [7], such as ProtoAI [14] and AILIXR [34]. This understanding helps to inform the development of the project and ensure its future success. In order to gain additional insights from the related product research, the exploration went beyond digital prototyping. It includes a variety of tools such as card kits and mindmaps. During the exploration process, the following three main design tools emerged that provided significant inspiration for the project.

The most outstanding one among these tools is the guidelines for HAI interaction, a collaborative effort involving Aether, Microsoft Research, and Office [20]. This guide holds a wealth of best practices and success stories, showing the desired behavior of AI systems across various scenarios, including initial interactions, routine engagements, problem–solving instances, and evolving interactions over time. It shows practical advice and corresponding success stories tailored to each of these contexts. Notably, it has a well–organized structure and easy to read. However, while the guide primarily addresses Human–AI Interaction, some of its recommendations, such as "Make it clear what the system can do," touch on issues relevant to traditional design. The extent to which these elements are specifically expounded upon in the context of AI remains somewhat unclear. See *Figure 4*.

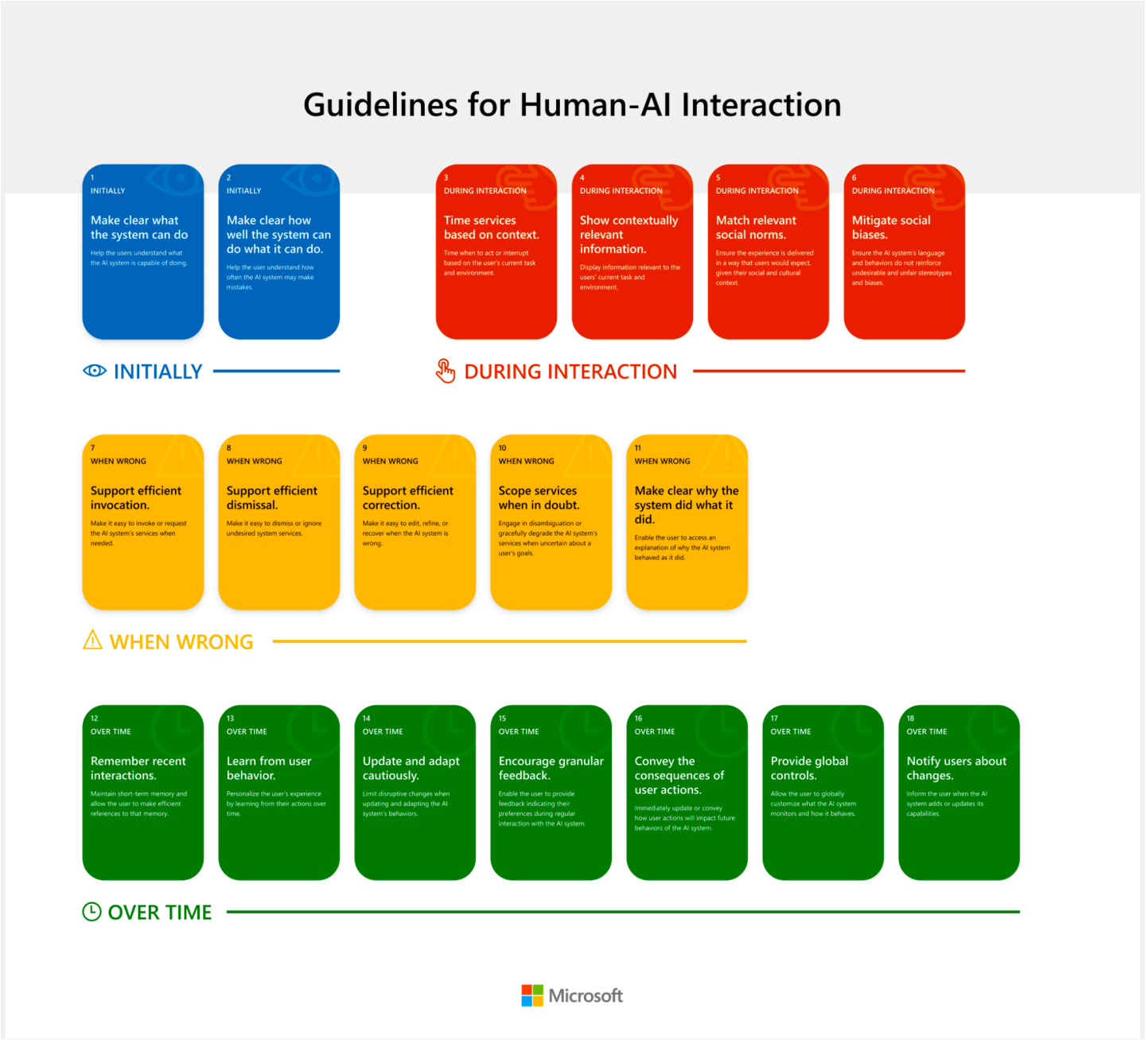


Figure 4. Guidelines for HAI. [20]

The second tool is Voiceflow [35]. this is one of the common digital tools on the market used to build Human–AI Interaction. Its main benefit lies in creating interactions between humans and voice assistants. And it support designers to review by providing the final design output like the UI interface. At the same time, it is also relatively simple to build, easy for designers to understand. But the interactions which can be created are limited, with few other choices. It is not sufficient for a wider range of design challenges. See *Figure 5*.

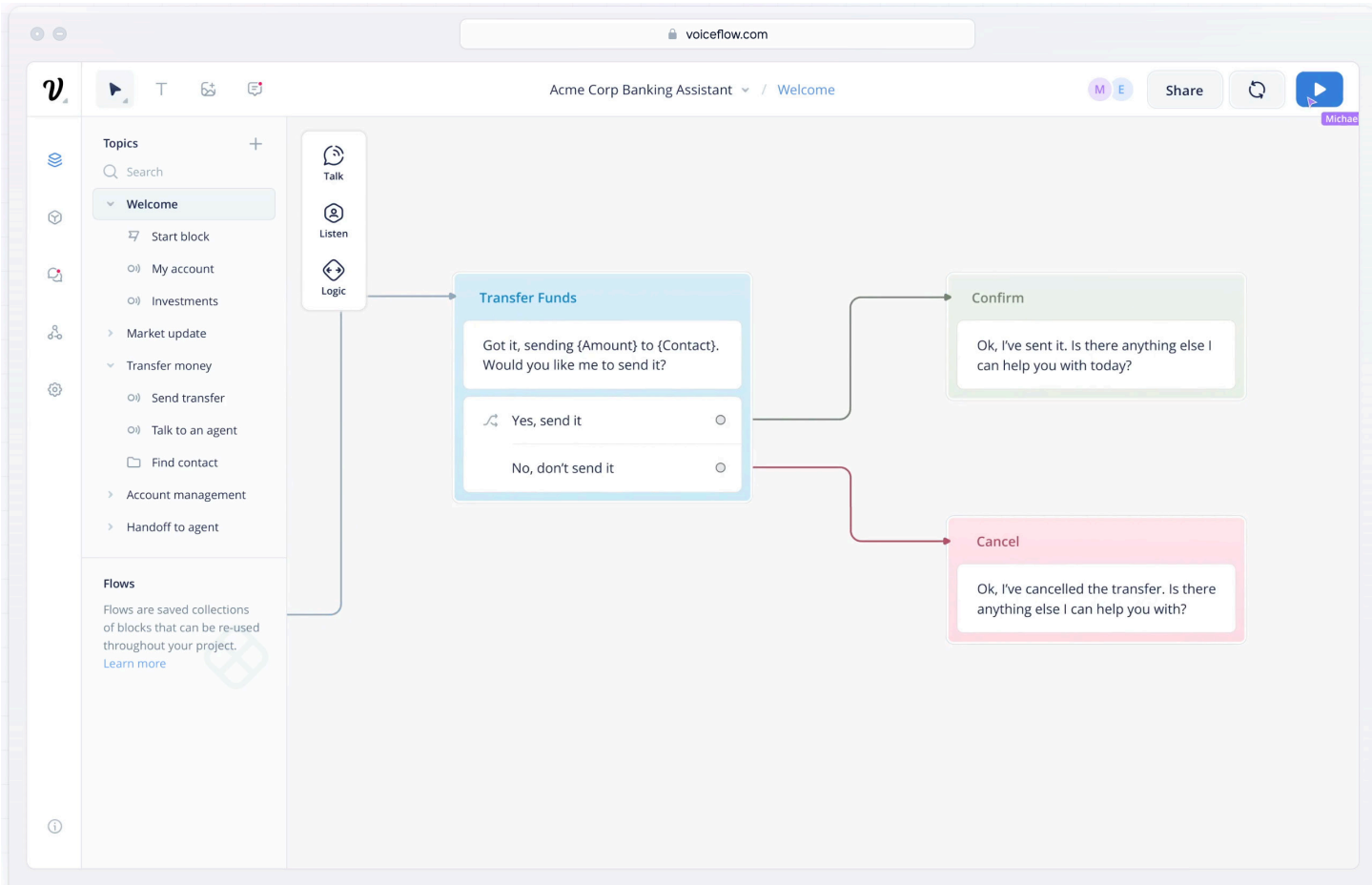


Figure 5. Overview of Voiceflow. [35]

The third tool is the AI meets Design toolkit. This was developed in collaboration with Nadia Piet [36] and Mobgen | Accenture Interactive Amsterdam in 2019 to collaborate on the development of a product that aims to help designers and innovators to design with machine intelligence at every step of the design (thinking) process. This toolkit is very detailed and inspiring to help designers in thinking about how to use AI as UX material. However, they do not help designers to produce the required interaction prototypes. See *Figure 6*.

Figure 6. Overview of the AI meets Design toolkit. [36]

2.6 Design considerations

Designers in this project consider three main aspects while designing HAI: the content related to the concept Message & Message sequence, the content related to instances in use cases and how to help end-users to understand AI models' output & feedback/XAI through the user interfaces which are presented on nature languages or visualization. As shown in the *Figure 7.1*.

In general, current research on design considerations has focused on various aspects, including UI interfaces [13, 20], initial design [12, 13, 14, 20], interaction levels [12, 13, 14, 20], error correction [12, 13, 20], and long-term operations [12, 13, 20]. Considering the characteristics of this project, we followed the design considerations for the model-informed prototyping [14]. Model-InformedPrototyping (MIP) is a workflow that combines model exploration and interface design tasks. See *Figure 7.2*.

Specifically, design considerations will have an impact on the functions and interactions of the final design outcome of this project, so it is important to rationalize design considerations from the literature with those provided by the client. This element will be explored in the Pre-phase in Chapter 3. And for the connection between the final design features and the specific design considerations for the project, see Chapter 4.

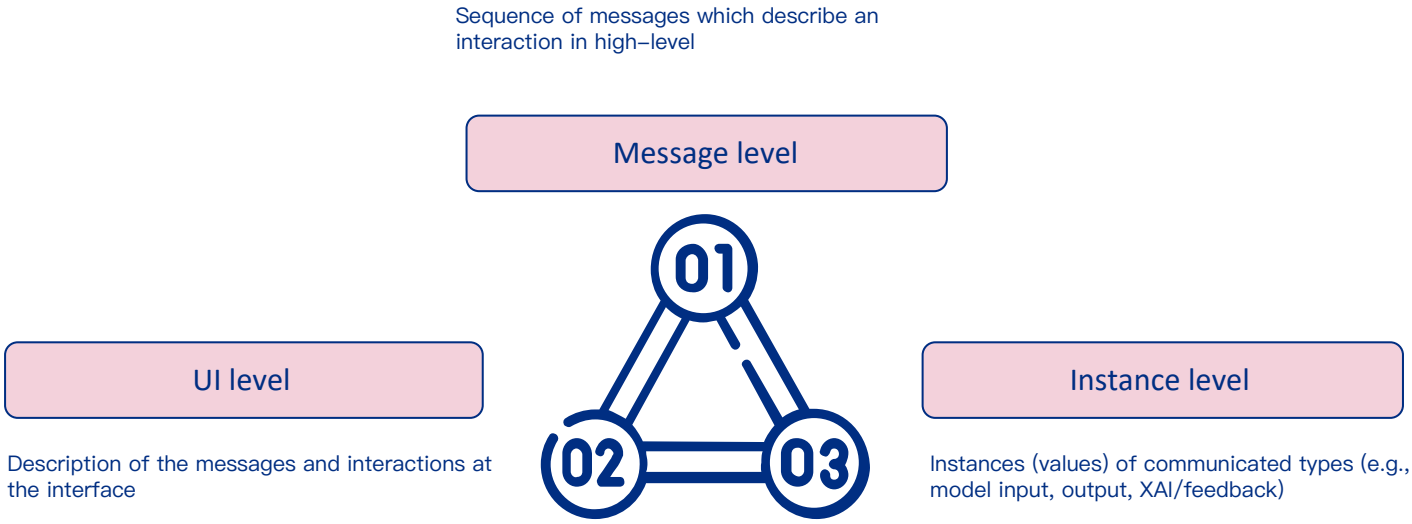


Figure 7.1. Material from client, but the more specific design considerations for this project derived from these are in Chapter 3 in the subsection related to Insights in Pre-phase.

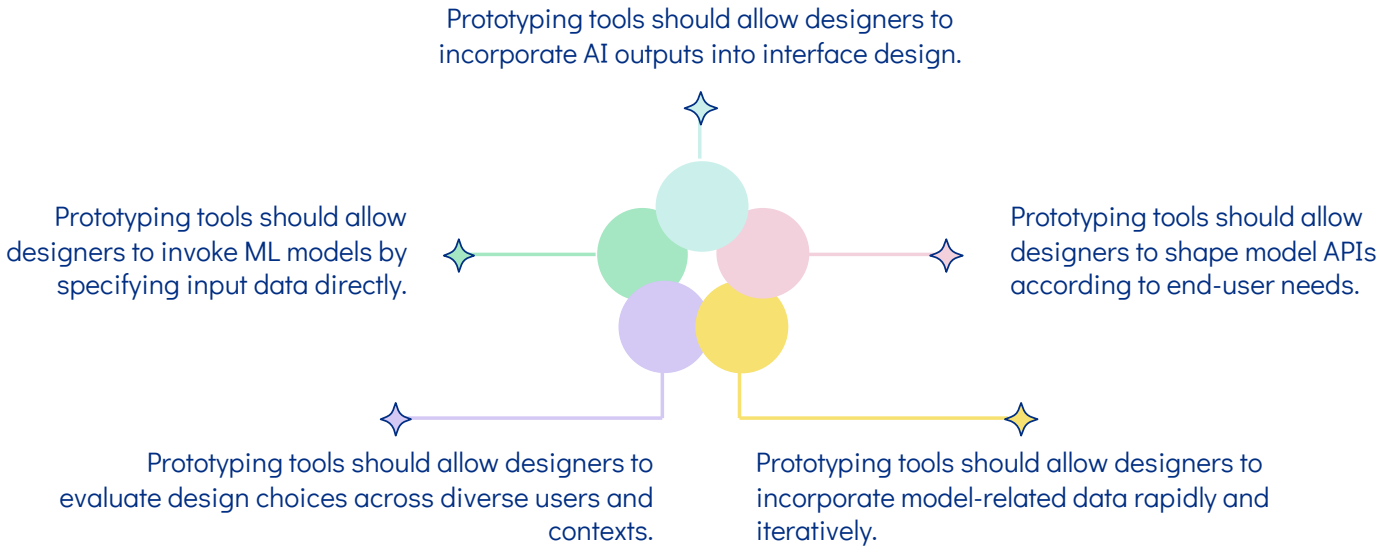


Figure 7.2. Design considerations for model-informed prototyping.[14]

3.

Test for research

The third section will show the two testing activities that occurred in Pre–Phase as well as in Phase 1 to practice the design methodology. Each activity will be presented with details on the preparation, the process, and the results or insights from each activity. This structured approach aims to provide a overal overview of the testing process and its outcomes.

3.1 Pre–Phase

It is seen from the previous section, basic concepts including the definitions of the 12 types of information, Message & Message sequence, and HAI were established at the beginning of the project, and a CV screening use case was created based on the content provided by the client (see Appendix C). The overall activities followed the process shown in *Figure 8*.

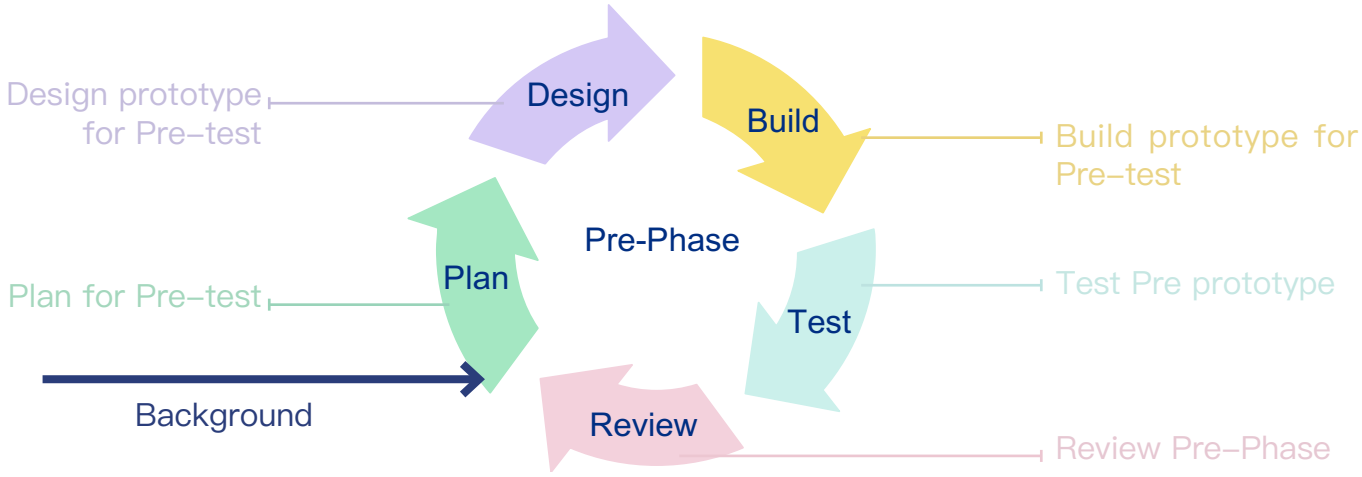


Figure 8. Overview of Pre–Phase.

As the project explored concepts like "Message" in AI design, it remained uncertain if designers could readily grasp and integrate these terms or tools into their design processes.

Meanwhile, based on previous analysis of Voiceflow and other mainstream AI/UX products, it can be seen that in most cases, maybe it is the designer's mindset, or maybe it is a problem with the product design process, the workflow that is ultimately designed for the end–users who will use the products is often a linear data structure.

Data structure where data elements are arranged sequentially or linearly where each and every element is attached to its previous and next adjacent is called a linear data structure, like tables or arrays [37]. Data structures where data elements are not arranged sequentially or linearly are called non–linear data structures., like Trees [37]. See *Figure 9.1 & 9.2*.

This prompts a crucial question: "Is there any possibility for other non-linear data structure for HAI?" Answering this question holds significant implications for future designs, influencing factors like basic workflow complexity.

Figure 9.1. Example of Linear structures

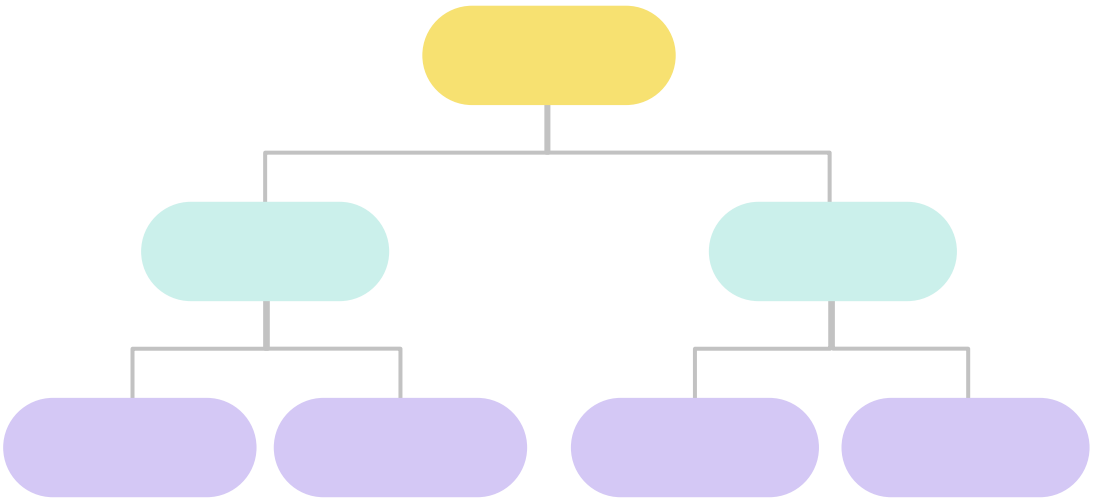


Figure 9.2. Example of Non-linear structures.

Also, as mentioned earlier, identifying design considerations would affect the functions and evaluation of the design outputs at a later stage. The earlier these considerations are established, the easier and more efficient the later work will be. Therefore, finding out how to combine the content provided by the client with the content in the literature was also one of the important goals of this test.

Research questions:

1. Can designers understand underlying concepts like Message and Message sequence? Or how to help designers understand underlying concepts like Message and Message sequence?
2. Is there potential for non-linear data structures when designers create Human-AI Interaction?
3. How to streamline design considerations specifically for the project by integrating design considerations from the model-informed prototyping [14]?

3.1.1 Preparation

The test in this phase is called Pre–test. To answer the Research questions mentioned, the test involved 7 steps (see *Figure 10*) and 3 activities. These activities were introduced step by step, aiming to help participants understand the concepts in the project. The main goal was to educate participants not only to understand these concepts but also to use these concepts effectively in creating interactions for specific the use case CV screening. This incremental approach was executed to facilitate a progressive and structured learning experience for the participants. For details about the test plan, see Appendix D.

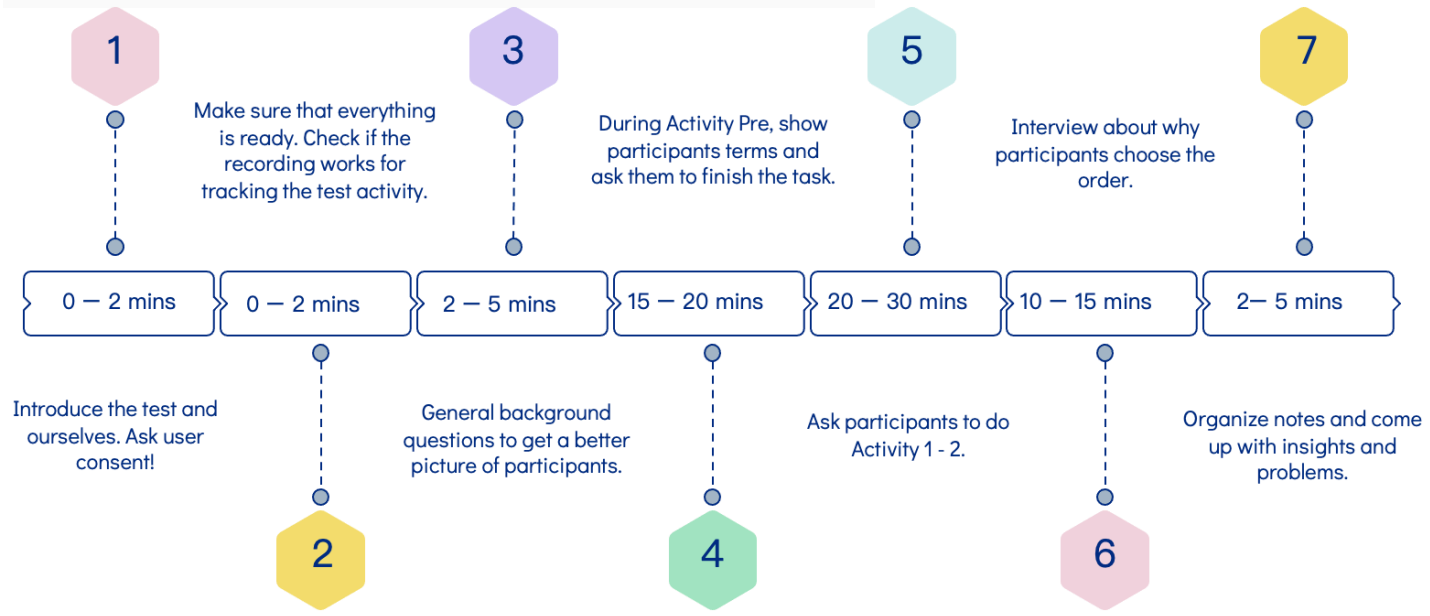


Figure 10. 7 steps for the Pre–test.

The following is a detailed description of the purpose of the three activities, their content, and how they contribute to research questions.

Research questions:

1. *Can designers understand underlying concepts like Message and Message sequence? Or how to help designers understand underlying concepts like Message and Message sequence?*

2. *Is there potential for non-linear data structures when designers create Human-AI Interaction?*

3. *How to streamline design considerations specifically for the project by integrating design considerations from the model-informed prototyping [14]?*

Activity-Pre

Goal: Evaluate participants' understanding of underlying concpets "Communicative acts" and HAI.
Overview: Researchers provide an overview of the project and definition on the underlying concepts. Participants familiarize themselves with Message, enhancing both terminology understanding and practical application in the project context.
Function: Acts as a litmus test for the effectiveness of the introduction, assessing participants' readiness for subsequent activities.

Activity 1

Goal: Strategically balances the overall learning difficulty after the introductory task.
Overview: Involves a more familiar task for designers, constructing the CV screening scenario within a specified use case.
Function: Support the research questions related to linear and nonlinear data structures. Enables participants to smoothly transition while applying their knowledge in a context aligned with their design expertise.

Activity 2

Goal: Supports the test objectives related to linear and nonlinear structures while focusing on participants' practical application of knowledge.
Overview: Participants fill in paper Message cards and create sequences for the context established in Activity 1.
Function: Support all the research questions. Enable Participants to show what they have learned and understood.

Considering that linear structures are probably the easiest to consider, some possible structures are provided in this step in order to give participants some inspiration. See *Figure 11*.

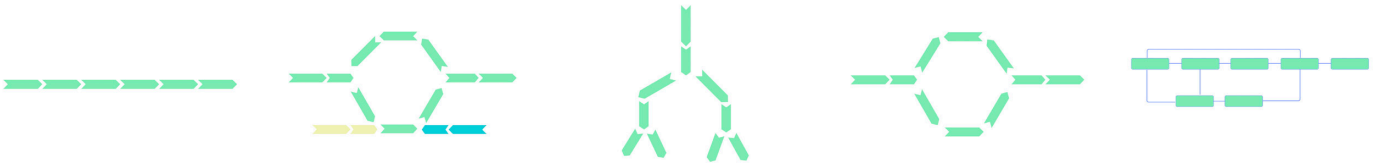


Figure 11. Possible structures to inspire participants.

For details about paper prototype and other materials using in the Pre-test, see Appendix D.

3.1.2 During the test

A total of 2 participants were invited to participate in this test, Design for Interaction and Strategic Product Design from TUD's IDE Institute. The files of the voice transcriptions from the test as well as the content they created will be shown in Appendix E.

Here is one figure showing what participants did during the tests in general. See *Figure 12*.

Overall, participants exhibited commendable progress in gaining a foundational understanding of the project's concepts upon the successful completion of the first activity. This first activity served as an important foundation, enabling participants to understand the basic concepts of the project. As they seamlessly transitioned to the second and third activities, participants demonstrated not only a retained understanding but also a capacity to apply their knowledge with a degree of autonomy. This adaptive competence showcased their ability to explore and engage with the later tasks in a style that reflected both their understanding and individual problem-solving approaches. The collected outcomes emphasized the effectiveness of the learning structure and participants' willing to use acquired knowledge in the context of the project.

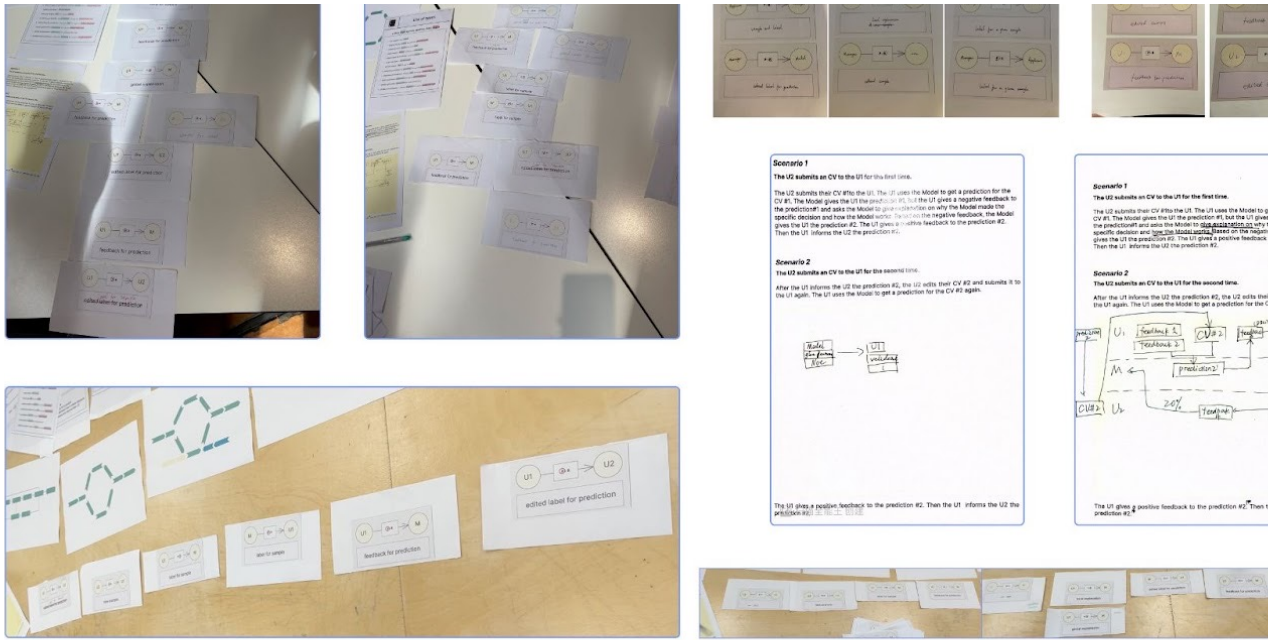


Figure 12. Participants in the test.

3.1.3 Insights

The analysis has led to 2 key insights, each holding important effect for the development of design concepts, and the development of design considerations.

For the concepts in the project, the participants both agreed that although it took a lot of time to understand them and there were many suggestions, these concepts made sense for them.

“ (After talking about difficulties and questions during the Activity Pre)The rest are easy for me. ” — Participant 1

The detailed explanation of these insights add depth to the understanding of how designers considered Human–AI Interaction, establishing a strong groundwork for continued exploration and practical application in upcoming design.

1. Insights on structures

While many factors may influence the arrangement and final outcome of Message, designers preferred a linear structure as the best worflow structure. See *Figure 13.1*. The reason is rooted in the efficiency of conveying information through a linear sequence.

Non–linear structures, viewed as supportive elements, are often considered derivative that do not impact the primary linear sequence a lot. That means although non–linear processes may be involved in some of the task flows, when viewed as a whole, they might not necessarily affect the final result, and the entire task flow still presents a linear structure. See *Figure 13.2*.

In conclusion despite the prevailing preferrence towards linear structures, discussions have arisen regarding the potential effect of opting for a linear or non–linear structure on the overall Human–AI Interaction (HAI).

"The story is every linear. And it is the fast way to solve problems." — Participant 1
"It's like a logic diagram, with "Yes" or "No" going to the next step or continuing the loop." — Participant 2

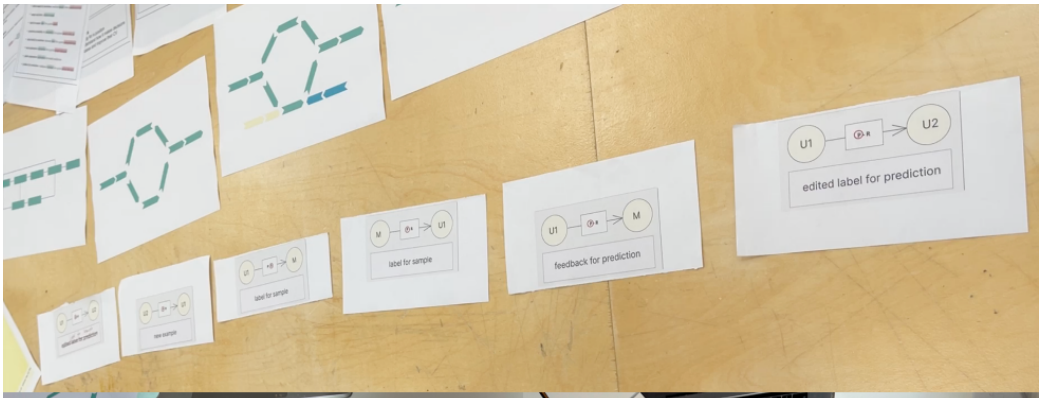


Figure 13.1.Example of the Linear data structure.

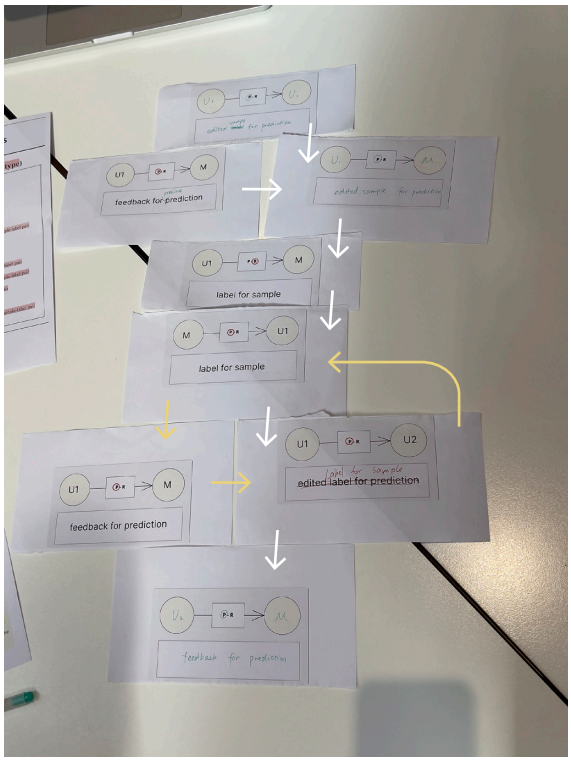


Figure 13.2. The overall presentation is linear, but there is a nonlinear data structure in it.

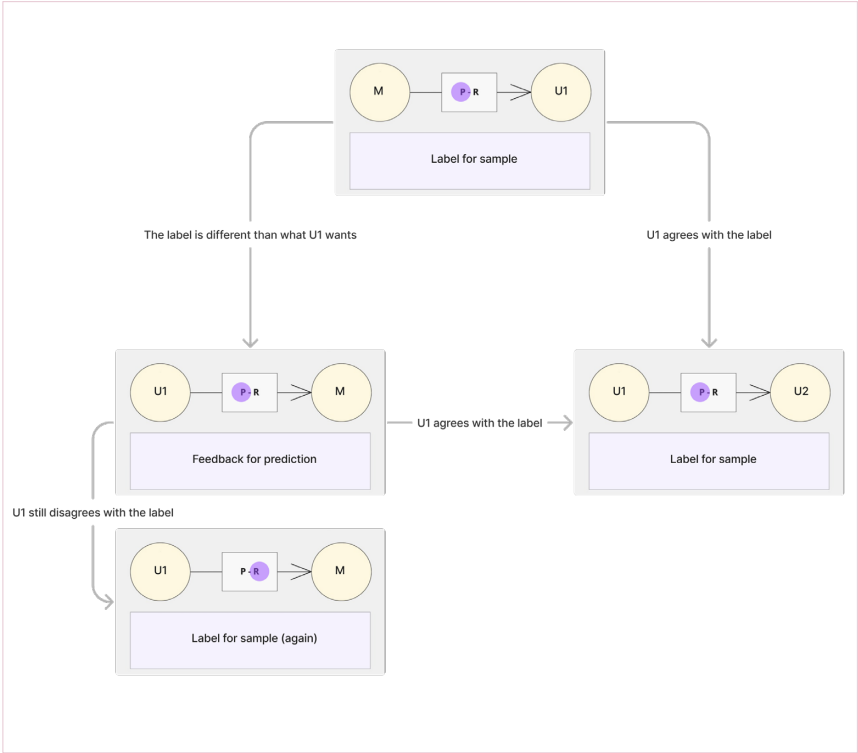


Figure 13.3. A part of the reorganization shown in Figure 13.2.

Figure 13.2 shows the message sequence created by Participant 2 based on the story created in Activity 1. The white arrows in it represent the interactions that Participant 2 believed occurred between the the target user (U1) and the AI model in the context in order to achieve U1's own goals. The yellow arrows represent HAIs that might be present, and sometimes they might affect the interaction represented by the white arrows. Based on Participant 2's words, one part can be organized as shown in *Figure 13.3*. These yellow arrows represent a non–linear structure. Participant 2 argued that when the AI provides a prediction that is different from what U1 envisions, U1 may create different choices after reading the explanation, and these choices also convey different data or information.

2. Insights on Design considerations

Each participant had their own way of thinking about HAI, but interestingly, a few similarities stood out. Both participants tended to focus on creating interactions that directly solved specific problems before adding other interactions to improve the overall user experience. This emphasis on solving problems was different from some participants who concentrated on creating interactions related to giving feedback. It can be seen a visual representation of these approaches in *Figure 14.1*.

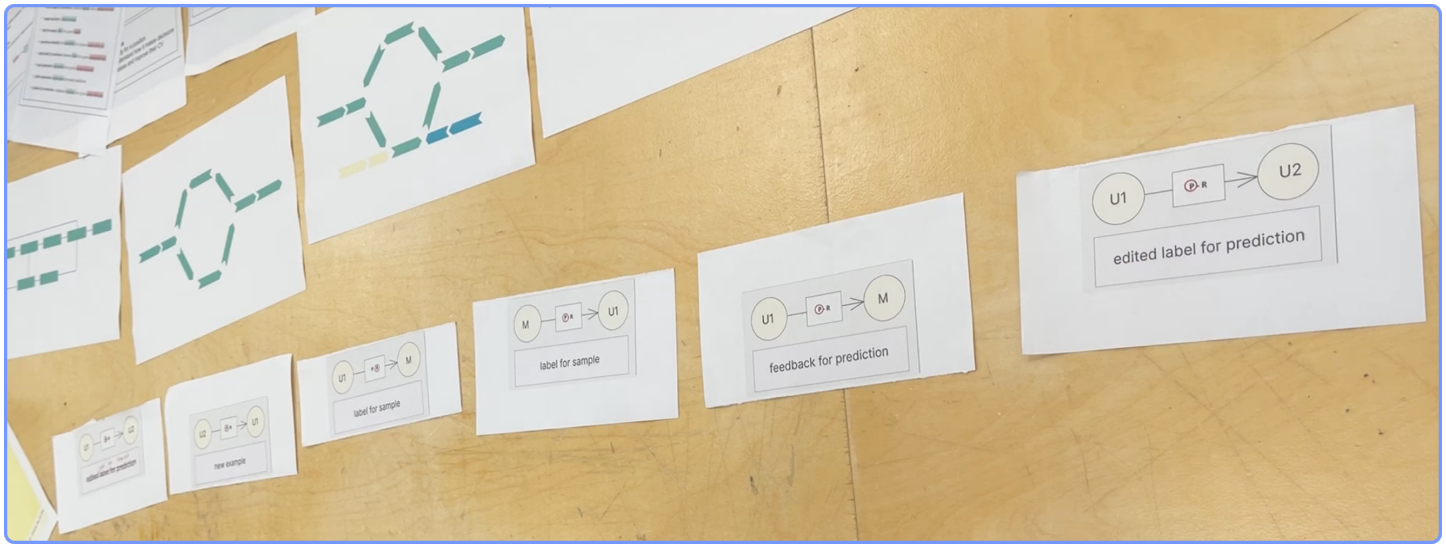


Figure 14.1. Build sequences for problem–solving interaction first.

Also when they were building interactions, they thought about how those interactions were presented. Is it a voice assistant? Is it communicated through a text–based dialog format? Or is it some other types of Graphihcal user interface? These thoughts affected the behavior of the interaction to some extent, but would not have much impact on the overall sequence of information. See *Figure 14.2*.

" While we have to think about the sequence on the local explanation, for me, it depends on the interface. " — Participant 1

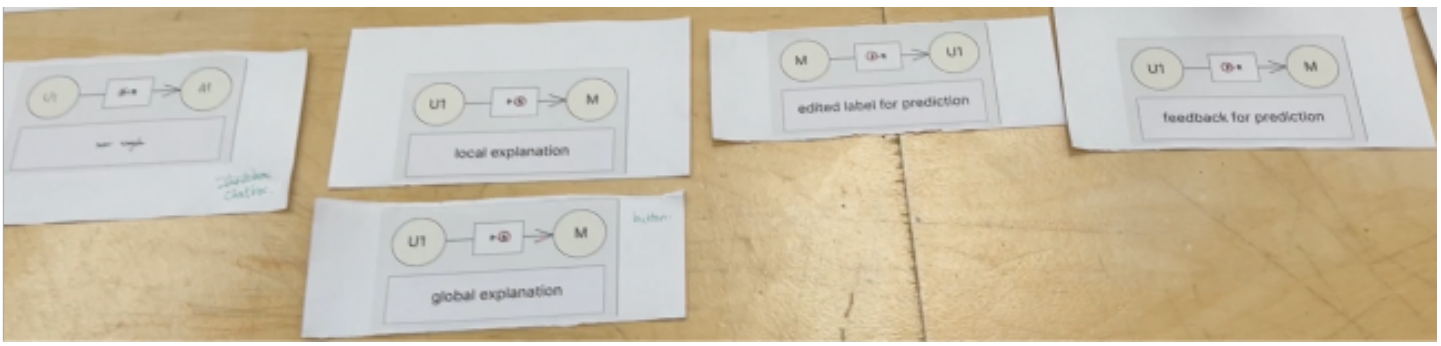


Figure 14.2. How to present interactions.

There was indeed a discussion about "decision–maker" for this section during the analysis, but ultimately it was deemed that this section was not a part of this project that should be focused on exploring. For more details in Appendix E.

Combining the insights brought by these participants with the previously mentioned theoretical basis (in Chapter 2), the particular design considerations of this project were summarized in the "Design Considerations for the project"as shown in *Figure 15*.

In *Figure 15*, it also shows possible functions based on the design considerations for the project. In Chapter 4, the final chosen functions in the design prototype would be discussed in details.

These design considerations also contribute to the subsequent design goals in Chapter 4. They emphasize the importance of the simplicity in the final design as well as the presentation of concepts related to "communicative acts". What's more, they also provide possible approaches.

Design Considerations for model-informed prototyping [13]	**Design Considerations for the project**	**Possible functions**
Prototyping tools should allow designers to invoke ML models by specifying input data directly.	1. Designers can create, modify, delete and move messages to form interactions. 2. They can modify the instances or data information about inputs/outputs/instances, etc. of the AI models according to the needs of the end–users for evaluating the predictions.	Create & modify messages & message sequences; choose the best AI models based on the end–users' needs; connect instances in the use case with messages; visualize the input data during the HAI…
Prototyping tools should allow designers to incorporate AI outputs into interface design.	Designers shall be able to visualize in the interface the instances of each Message that are related to the exchange of information or the HAI–related UI elements.	Visualize the output & feedback/XAI; connect instances in the use case with messages; choose the best AI models based on the end–users' needs…
Prototyping tools should allow designers to shape model APIs according to end–user needs.	Designers can define the inputs/outputs/Feedback–XAI of the AI model and the presentation of these based on the needs of the end user in the use case. However, they should not involve too much coding and focus more on how the designer builds the HAI.	Choose the best AI models based on the end–users' needs; help designers understand functions of each AI model easily…
Prototyping tools should allow designers to evaluate design choices across diverse users and contexts.	1. Designers can create personas that will perform HAI based on use cases, whether they are humans or AI models. 2. The design outputs should be broadly applicable to different design challenges.	Create & modify messages & message sequences; connect the end–users' needs with the AI models and messages…
Prototyping tools should allow designers to incorporate model–related data rapidly and iteratively.	Designers should have the flexibility to adapt the content created in the design output to the needs and feedback of the end user.	Create & modify messages & message sequences; visualize the data or instances conveying during the message sequence…

Figure 15. The project's design considerations.

3.2 Phase 1

With the Pre-test completed, the project provides initial insights into how designers can use Message to build Message sequences (HAIs) and provides design considerations. These insights are important guidance for future designs and deserve careful study.

However, there were still 3 questions waiting for answers. Therefore, in Phase 1, the research questions will be explored through Test 1. The flow of activities in Phase 1 is shown in *Figure 16*.

Research questions:

The exploration of creating a logical workflow in a digital interface and addressing the design considerations outlined above raises the following questions:

1. How can a Message containing communicative acts be effectively presented in the digital prototype?
2. What strategies can be used to represent a message sequence in the digital prototype?
3. How to apply the specific project's design considerations for the digital prototype?

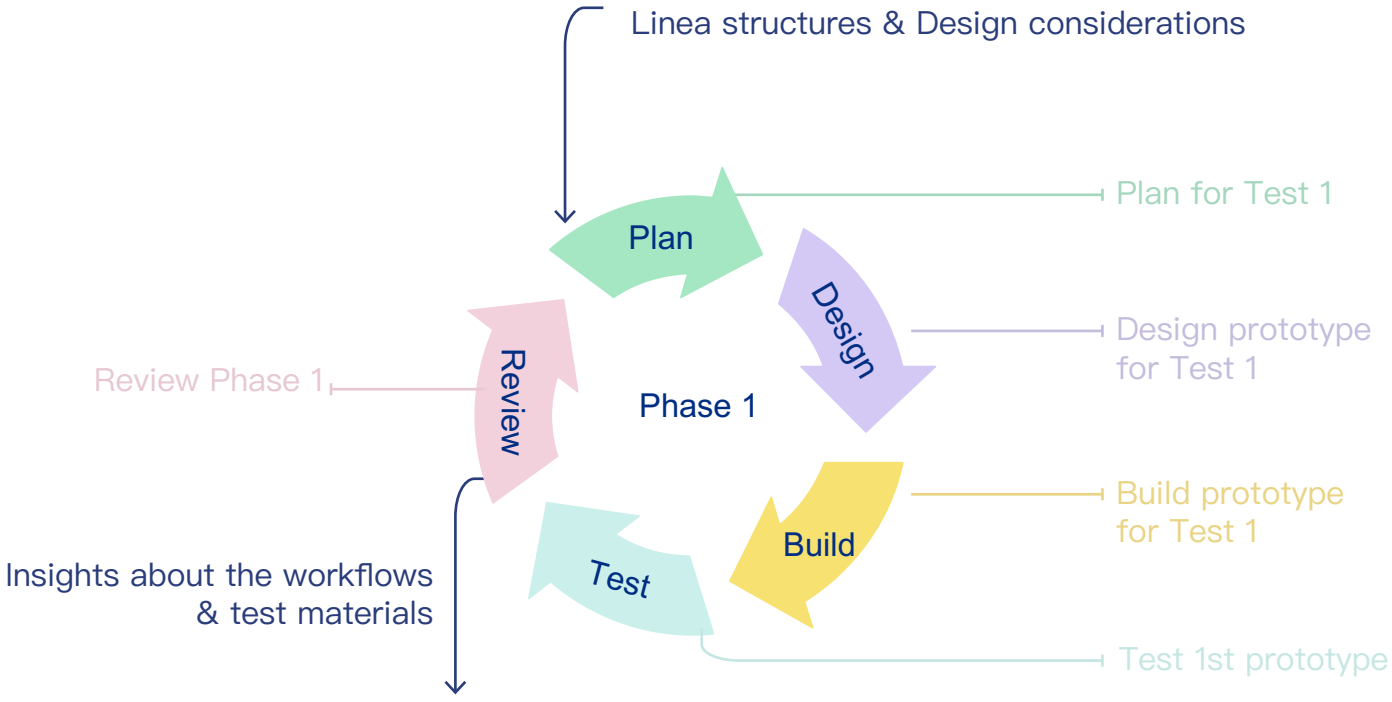


Figure 16. Overview of activities in Phase 1.

3.2.1 Preparation

To fulfill these objectives, the test held a total of 7 steps (*Figure 17*) and 2 activities. Building on the step–by–step approach used in the Pre–test, these activities were crafted to facilitate participants' understanding of new concepts related to digital prototyping, including elements like message, sample, label, feedback, etc. The general goal is for participants not only to understand these underlying concepts but also to proficiently apply them to interactions aligned with the provided use cases.

It is important to note that the main focus of this test is not to evaluate the functionality of the digital prototypes themselves but rather to investigate whether the digital prototypes effectively aid participants in grasping the concepts embedded in the project. Hence, in the introductory, researchers provided an overview of the basic functions of the digital prototype, explaining the contents of different areas and explaining how the digital prototype can be interacted with (e.g., dragged, clicked, added, etc.).

The final objective was to empower participants to not only acquire a solid understanding of these new concepts and streamline their learning curve in digital prototyping but also to proficiently apply these concepts to interactions relevant to their assigned use cases.

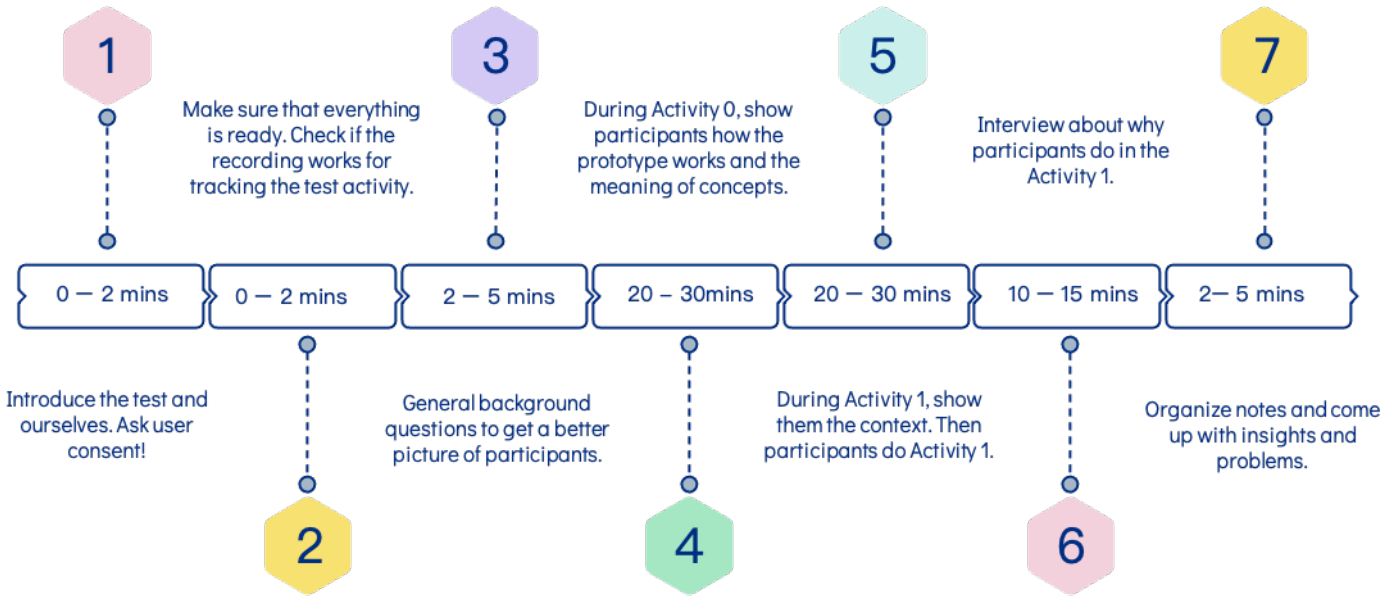


Figure 17. 7 steps for the Pre–test.

The following is a detailed description of the purpose of the 2 activities, their content, and how they contribute to research questions. For more details about activities and materials, see Appendix F. For the use case using in the test, see Appendix C.

Research questions:

1. How can a Message containing communicative acts be effectively presented in the digital prototype?
2. What strategies can be used to represent a message sequence in the digital prototype?
3. How to apply the specific project's design considerations for the digital prototype?

Activity 0

Goal: Familiarizes and educate participants underlying concepts in the project with a low-fidelity prototype's functionality.
Overview: The researcher provides a project overview using a low-fidelity model, explaining the connections between Message, Communicative acts, Message sequences, and HAIs. Participants create a new message aligned with the provided context.
Function: Acts as a litmus test for the effectiveness of the introduction, assessing participants' readiness for subsequent activities.

Activity 1

Goal: Evaluate participants' understanding of using Message to create HAIs in the low-fi digital prototype.
Overview: Participants showcase their usage of Message in the low-fi prototype to generate HAIs. They present instances related to the use case and identify logical UI elements for each Message.
Function: Collect participants' actions and statements to determine if the prototype aligns with design considerations and meets their needs. Collect data to assess the logical approach for designers to create HAIs using the digital prototype.

3.2.2 During the test

A total of 6 participants were invited to take part in the test: Design for Interaction, Integrated Product Design, and Strategic Product Design from the TUD IDE Institute. The speech transcription files from the test and the content they created are shown in Appendix G.

Here is one figure showing the message sequences participants did during the tests in general. See *Figure 18*.

In summary, participants showed commendable progress in understanding fundamental project concepts after observing the researchers' demonstration using the low-fidelity prototype. This hands-on experience significantly contributed to their foundational knowledge.

Moving on to the activities in Activity 1, participants displayed notable proficiency in creating Message cards. They skillfully assigned instances to different pieces of message, showcasing a careful understanding of the project's complexicity. Moreover, participants demonstrated their understanding by crafting diverse instances of the user interface (UI), each thoughtfully designed based on their understanding of the underlying concepts. This iterative engagement underscored the effectiveness of the learning process and the successful application of acquired knowledge in practical design task.

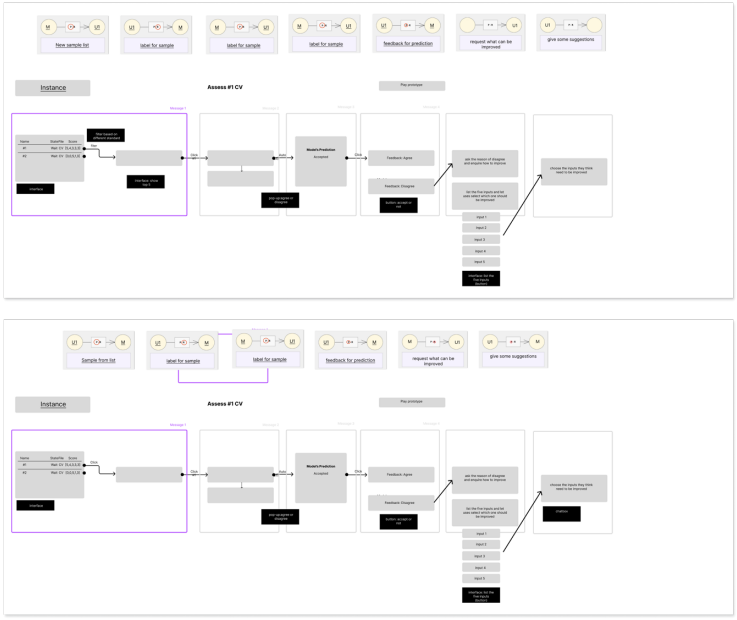


Figure 18. Examples on the message sequences which participants created during the test.

3.2.3 Insights

Although the number of participants in this testing phase was limited, the information gathered proved to be marked and insightful. The analyzingl approach involved brainstorming and summarizing the information extracted from the transcriptiont. The transcription and the process of analysis is depicted in Appendix G below.

This analysis highlighted 3 key findings about design and 1 insight on test–setup, each of which has important effect for conceptual introductions, interface workflows, and information hierarchies. These findings contributed to the subsequent development of the design output.

1. Emphasize and clearly demonstrate the connection between Instance and Terms.

Instance comes from the design considerations provided by the client in Chapter 2, and represents specific items associated with the use case, such as CVs, etc. Terms here specifically refer to sample, label/prediction, explanation, etc.

The flow starting Human–AI Interaction (HAI) design with the “Message" level seems to conflict with designers' natural instincts. Designers prefer beginning by defining instances within the HAI and outlining their attributes before creating sequential information cards.

This preference stems from their interaction with the design tool, where working on design activities within the instances section felt more practical and concrete. In contrast, the design tool developed based on Communicative acts in this study introduced a more abstract and conceptual workflow. As a result, participants using the provided tool had to invest significant mental effort in translating the tangible design content into the abstract content needed to test the prototype.

Unluckily, after navigating this cognitive process, participants found themselves compelled to shift back from abstraction to concretization. This cyclical process adds an extra layer of complexity and cognitive workload. As shown in *Figure 19*, participants transformed intsances in the use case into Messages (an abstract level) and ordered them. Now they were selecting the corresponding interactions and instances for visualizing HAI.

"So when the designers create the interaction way for the system, are the labels sure? I mean, do we have known what kind of labels we need?" — Participant 3
"I think Message is too abstract. When I need to create connection and UI components, I feel relaxed." — Participant 2

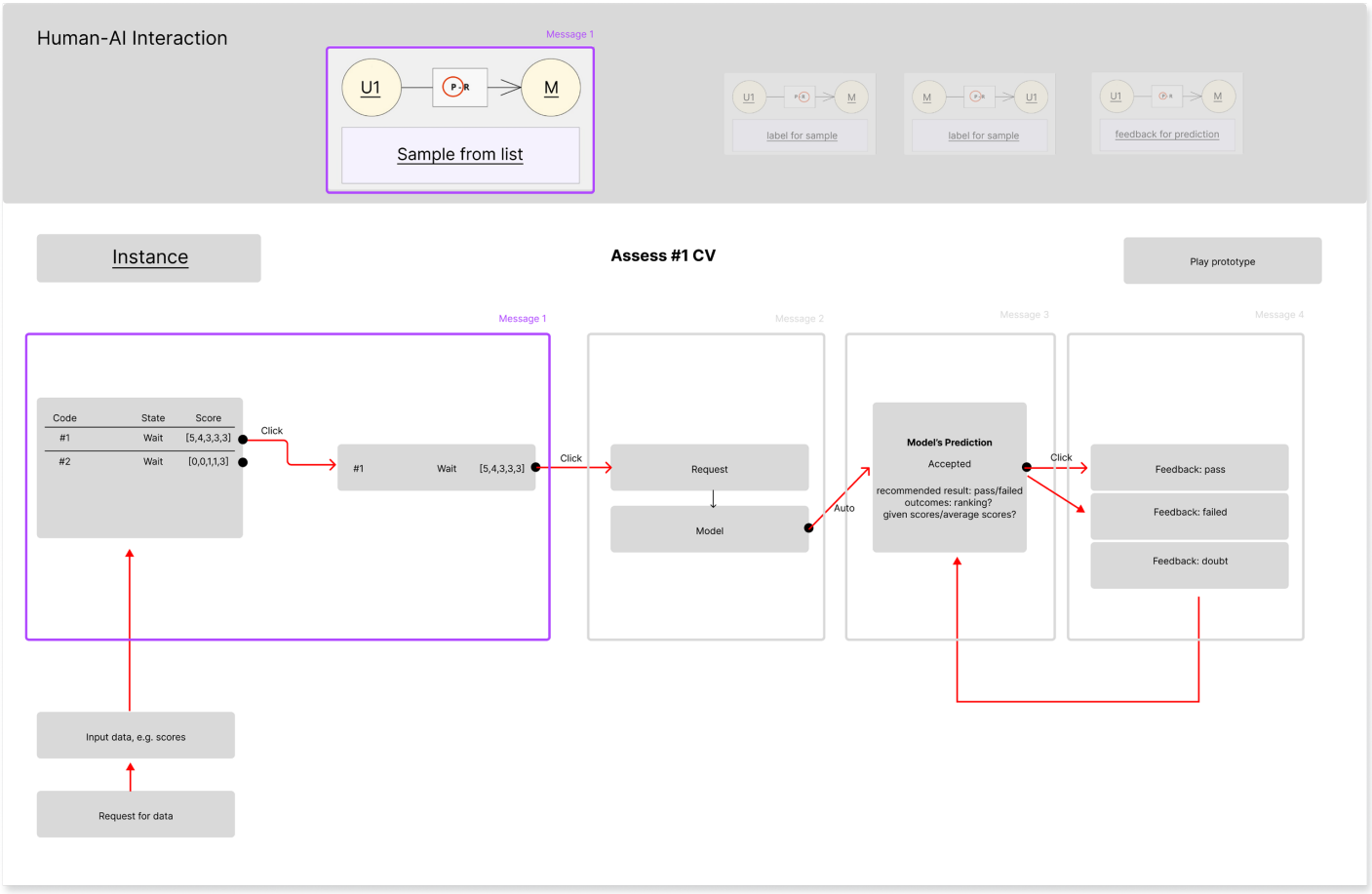


Figure 19. Materializing from Message level to Instance level & UI level.

This insight supports the idea that starting the prototype's definition of specific terms at the instance level would be more effective. In simpler terms, explaining terms like sample, label, feedback, etc., using instances could make things clearer.

Currently, information types are conveyed through phrases, but this method may not be optimal. Describing all information types as phrases can be confusing for designers when trying to define specific terms. Typically, participants prefer dealing with examples before getting into digital prototypes. So, when asked to come up with their own generalizations to identify the types of information they need, designers face significant challenges and a steep learning curve.

From a project standpoint, it might not be necessary to strictly define terms precisely based on examples. Instead, recognizing designers' reliance on examples for understanding, incorporating visual displays could be a helpful reference, making it easier to grasp different terms.

"How can we understand the sample, the label and the prediction? Please show me some examples instead of academic terms." — Participant 1
"What's the meaning of sample? Could you show me examples about it? And if possible, also the meaning of prediction, label, probability, explanation and feedback." — Participant 3
"What would the result be? I mean the prediction, how does it look like?" — Participant 6

Although participants suggested that it would be the most appropriate to start the workflow of the digital prototype with the "definition of terms using Instance", the intercepted quotes suggest that **the deeper need is for a way for designers to more clearly understand, learn, and see the relationship between terms and instances.** relationships with instances. This can be accomplished not only by making the terminology more relevant to the instances.

At the same time, a more abstract vocabulary may, to some extent, help designers think outside the box so that they can explore more HAI possibilities.

2. Fewer interfaces would be better

The fewer interfaces there are, the better designers might work. For the low-fi prototype, there were 3 interfaces and participants jumped between these interfaces a lot to build the UI elements. What's more, sometimes they would repeat the same behaviors again and again in different interfaces. For example, participants would think about how to create, view, and edit messages and message sequences, but also the UI elements *Figure 20*.

"I want to create instance/UI elements on message level interfaces. The fewer interfaces, the better." — Participant 1
"For the UI Level, I think it would be convenient to edit it in the Instance level." — Participant 1

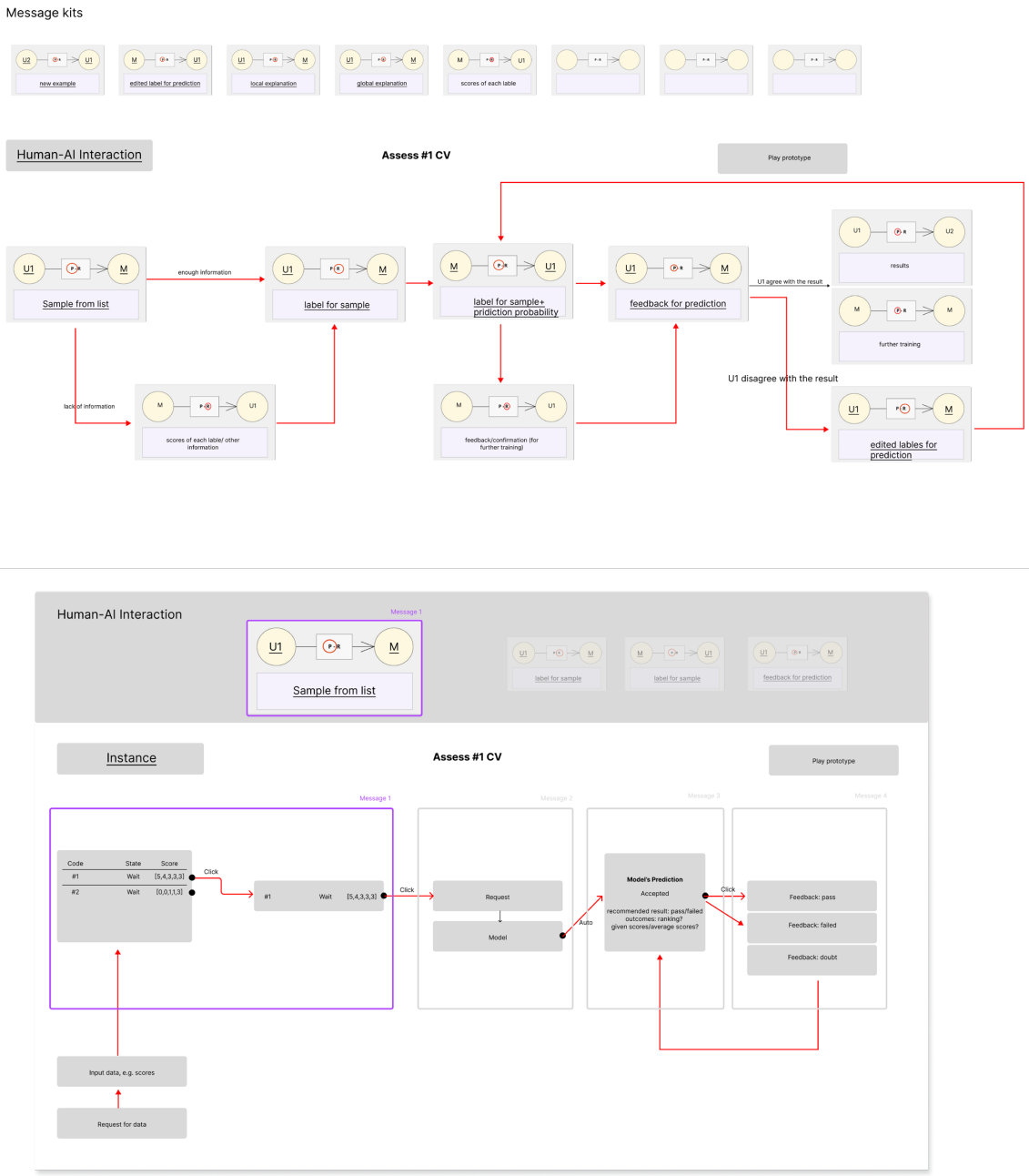


Figure 20. All participants prefer to create content from different levels in the same interface

This is reasonable because each interface jump interrupts the participant's design thinking to some extent. The more jumps there are, the more interruptions there will be. After the thought process is interrupted, participants need to spend time re-collecting and re-constructing based on it.

As a result, when participants used existing digital prototypes, their mental work could be tasking. Specifically, when they jumped to creating UI elements or instances after creating and arranging messages, they needed to spend some time reconstructing what they had already accomplished in the message level and rethinking the tasks at the existing level on top of that.

Reducing the jumps in the task interface can help designers think more smoothly along the lines of the original design and use prototypes.

3. Be careful to clearly distinguish and present content between different levels in the prototype.

It would be better if designers could have some freedom to create new items like terms/types of information by themselves. This is because sometimes designers might think the provided materials cannot meet their needs. Maybe they could have some freedom to edit these items by themselves. For example, in *Figure 21,* participants didn't find the right type of information to show the difference between different types of feedback. As a result, they created one type of information related to it.

However, in terms of the content created by the participants, the content they referred to as "not represented by the 12 types of information" (*Figure 21*) was actually more at the level of UI elements or Instance, etc. A reasonable explanation for this is that although some participants were able to complete the task, the existing concepts were still a barrier to their understanding. For example, the inability to distill the user interface information they need into more abstract terms, etc.

Another explanation lies in the fact that this reflects the importance of making detailed distinctions between functions at different considerations. Even if they have already understood and grasped the concepts in the project within a short period of time, the current digital prototype does not have a clear enough distinction between the different levels of content, which prevents them from properly establishing the correct connection between the abstract concepts in the Type of informations and the instances in the use case.

By improving the content presented in the second explanation, it is also possible to solve the problem in the first explanation: when the distinction between the individual consideration is clear, participants are also able to understand the underlying concepts more quickly and easily.

"I don't think modification and feedback for the more underlying logic of AI models (Feedback/ confirmation for further training)is the same as so–called Feedback–XAI. It's more feedback on the data trained or the class of model used." — Participant 1 (Figure 21)
"And I don't think there's anything (Request waht can be improved/Give some suggestions) in there that accurately describes this type of information. " — Participant 2

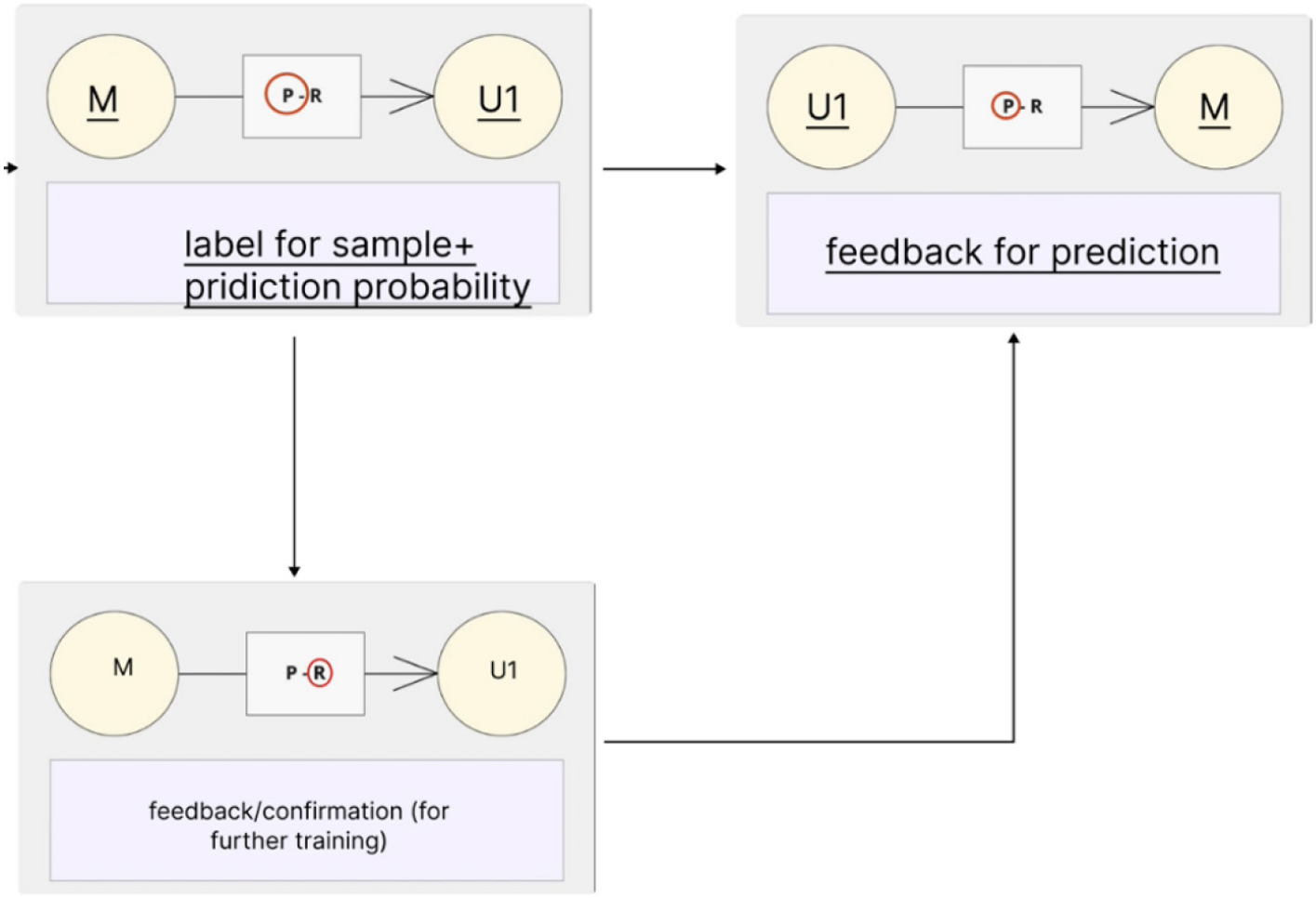


Figure 21. Example of what participants created by themselves.

4. Insights on test-setup

Participants not only shared the 4 main insights mentioned earlier but also offered feedback on the test, primarily focusing on suggestions related to the test's context.

The feedback highlighted a noticeable gap between the current context, the function of the AI model provided, and the real use cases. This gap led to certain tasks or interactions being unclear and challenging for participants to understand fully. Consequently, participants faced difficulties completing some parts of the content and providing meaningful suggestions.

To address this issue, it is important to enhance the test context and materials in subsequent tests. Providing a more reasonable test context and supporting materials will contribute to a better understanding, engagement, and participation of participants in future testing sessions.

"I think I need to know the motivation for the manager to use the model. You see, if the manager just uses the model to assess 100 or more CVs and then get the top 5%, it needs one interaction flow, However, if the manager wants the model to compare #1 and #2, there are another interactions. " — Participant 1
"What's more, I think it would be better if we could set one standard to assess the scores."— Participant 2
"To be honest, I think the context is a little confusing. For me, it would be more reasonable if it is the model chooses the CV from the list." —Participant 3

3.3 Conclusion

In this section, it went through Pre–Phase and Phase 1. During test activities in both phases, it obtained design considerations *(Figure 15)*, design suggestions and test improvement suggestions that guided the subsequent design.

Design suggestions

1. The linear structure is preferred. (Pre–Phase)
2. Emphasize and clearly demonstrate the connection between Instance and Terms. (Phase 1)
3. Fewer interfaces would be better. (Phase 1)
4. Be careful to clearly distinguish and present content between different levels in the prototype. (Phase 1)

Test improvement suggestions

1. Providing a more reasonable test context and supporting materials in the subsequent tests. (Phase 1)

Design Considerations for model-informed prototyping [13]	Design Considerations for the project	Possible functions
Prototyping tools should allow designers to invoke ML models by specifying input data directly.	1. Designers can create, modify, delete and move messages to form interactions. 2. They can modify the instances or data information about inputs/outputs/instances, etc. of the AI models according to the needs of the end–users for evaluating the predictions.	Create & modify messages & message sequences; choose the best AI models based on the end–users' needs; connect instances in the use case with messages; visualize the input data during the HAI…
Prototyping tools should allow designers to incorporate AI outputs into interface design.	Designers shall be able to visualize in the interface the instances of each Message that are related to the exchange of information or the HAI–related UI elements.	Visualize the output & feedback/XAI; connect instances in the use case with messages; choose the best AI models based on the end–users' needs…
Prototyping tools should allow designers to shape model APIs according to end–user needs.	Designers can define the inputs/outputs/Feedback–XAI of the AI model and the presentation of these based on the needs of the end user in the use case. However, they should not involve too much coding and focus more on how the designer builds the HAI.	Choose the best AI models based on the end–users' needs; help designers understand functions of each AI model easily…
Prototyping tools should allow designers to evaluate design choices across diverse users and contexts.	1. Designers can create personas that will perform HAI based on use cases, whether they are humans or AI models. 2. The design outputs should be broadly applicable to different design challenges.	Create & modify messages & message sequences; connect the end–users' needs with the AI models and messages…
Prototyping tools should allow designers to incorporate model–related data rapidly and iteratively.	Designers should have the flexibility to adapt the content created in the design output to the needs and feedback of the end user.	Create & modify messages & message sequences; visualize the data or instances conveying during the message sequence…

Figure 15. The project's design considerations.

4.
Ideas

The fourth section will introduce the current design goal, functions, and prototypes of different ideas made by Figma.

4.1 Design brief

This section explains the efforts and decisions made by the study to reach the final design output. The overall concept was developed based on the project context and the analysis and conceptualization of the insights tested in the previous phase. All choices were made on the basis of project research and user testing.

4.1.1 Design goals

Integrating the project background with insights from the prior testing phase, the design scope for this project was clarified as follows:

> My design goal is to design a **digital tool** that allows designers to prototype **Human–AI interaction** based on **communicative acts** through a **Human–center design process**. What's more, this tool:
> – is easy to learn how to use the tool
> – Help designers learn the basic concept "Communicative acts" of Human–AI Interaction involved in this project.

In addition, there are a number of sub–level design objectives that are also important but were not examined in depth in this project.

It's important to note that this tool doesn't help designers create front–end–like interfaces but only serves as a tool to inspire ideas on how to build an HAI. See *Figure 22.1*. Based on findings from previous testing and existing literature, the tool aligns with the Design Thinking [32, 33] by working as designers transition from the Problem Definition phase to the Design Ideation phase (i.e., phases 2 and 3 of design) and persists through the completion of phase 4, "prototyping."

What's more, while this design output is intended to contribute to a wider range of interaction design approaches (e.g., interacting with physical products) in the future, so far the focus has been more on helping designers build Human–computer interaction–related interactions such as interfaces [18]. See *Figure 22.2*.

At the same time, design considerations and design goals have different roles; design goals are what the overall design output is intended to achieve, and they are a more refined set of goals for the final outputs expected in the backgound. Design considerations, on the other hand, are goals for various aspects of the design outputs under design goals. Simply put, design outputs need to achieve the goals set by design goals by achieving the goals in design considerations.

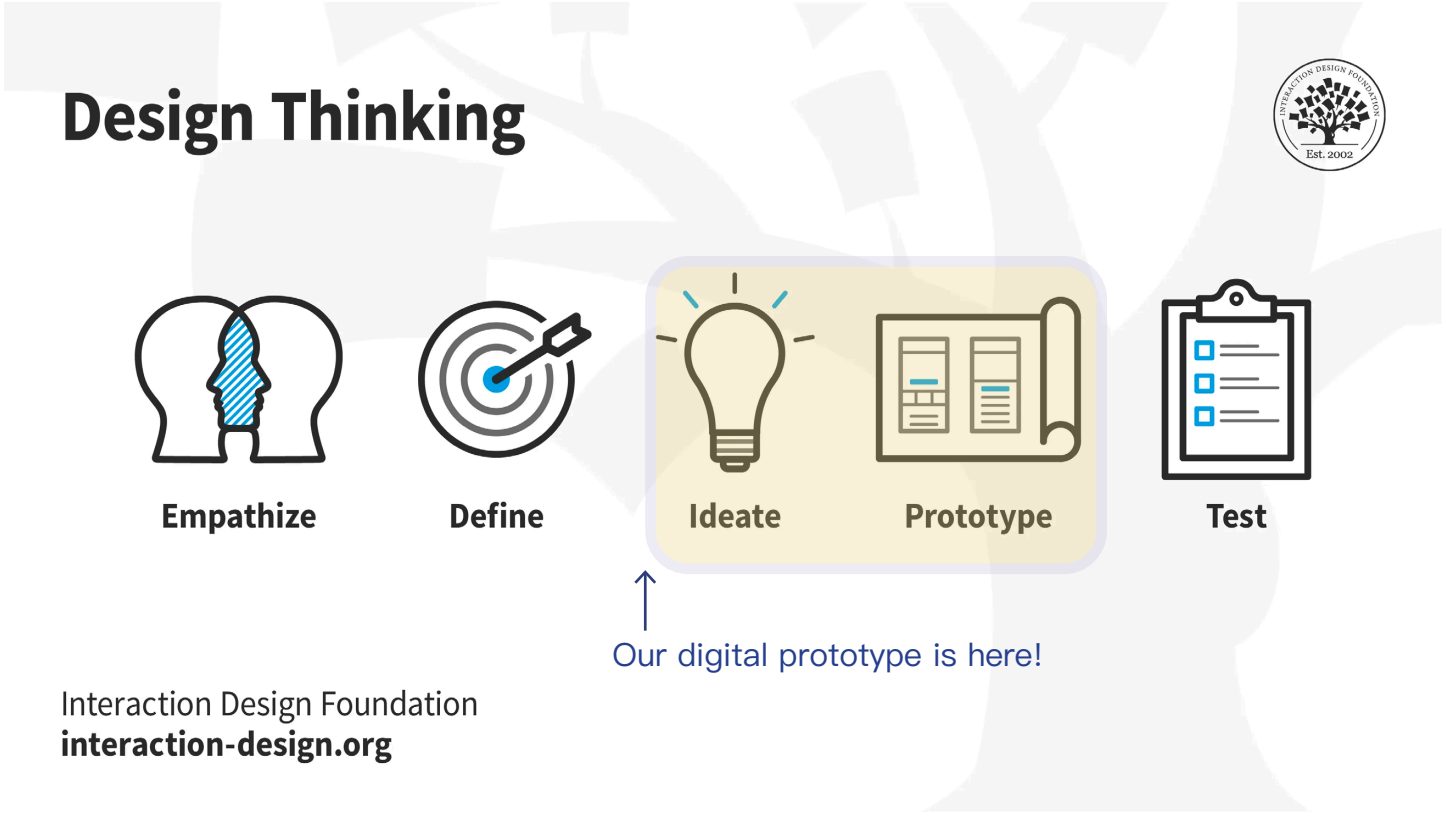


Figure 22.1. When to use the digital tool based on the Design thinking process.

Figure 22.2. Scopes of the project design output.

4.1.2 Information Architecture

Prior to advancing with the design, it is imperative to establish a consistent definition of the terms about the project and show the hierarchy of information among these terms. It will influence the organization of information across various locations during the design process and guide the selection of an appropriate design strategy.

Figure 23.1 shows the information architecture relationship between the terms that would be involved in the later design concept, and *Figure 23.2* shows the definitions of these terms.

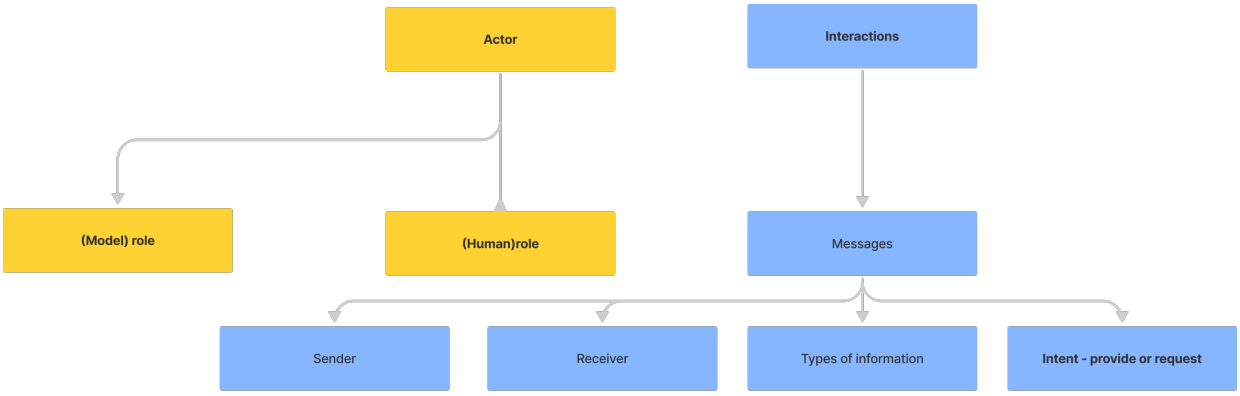


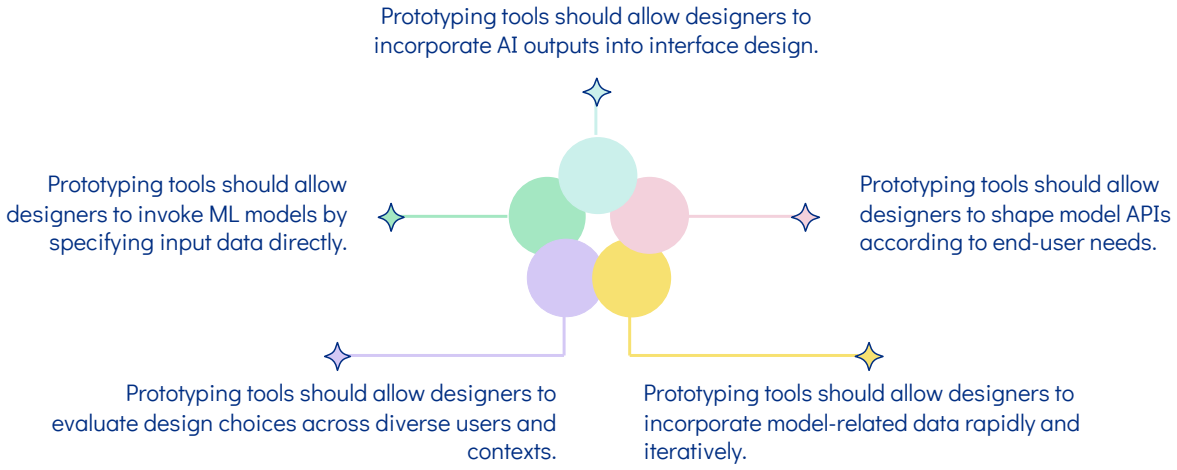
Figure 23.1. The Information Architecture of terms.

Term	Explanation
Actor	Humans or AI models who would be included in the HAI.
(Model) Role	Human readable name for this particular model, including functions or explanations in detail.
(Human) Role	Information about humans, like a persona.
Interactions	Content to craft Human-AI interactions.
Message	One single communicative act. In it, the sender communicates with the receiver by either providing or requesting a specific type of information.
Sender	Choice from one of the actors defined in the interaction-could be a person or a model.
Receiver	Choice from one of the actors defined in the interaction-could be a person or a model.
Types of information	Information which would be exchanged during the HAI.
Intent - provide or request	Actions for the Sender in one message.

Figure 23.2. The definiton of terms.

4.1.3 Features

To align with the project goal and design considerations, the interface should have the following functions. The 5 design considerations (in *Figure 7.2*) from the literature are here to succinctly represent the specific design considerations gathered from the Pre-test for this project (in *Figure 15*).



1. Create and show actors, including model and human roles

This one feature will require the creation of different roles depending on the needs of the design use case. Also, this is the essential Sender and Receiver part of building a Message.

2. Create and show new scenarios

This feature is designed to differentiate between different design tasks. When a designer is working, there may be several different contexts, and different contexts will have different HAIs.

3. Create and show messages

This function is for creating blank messages.

4. Create and show interactions (Connect messages)

This feature consists of composing sequences using messages and can change the order and direction of information flow on demand.

5. Help designers to learn the tool/terms in the tool

This feature demonstrates the need to consider how to reduce the learning costs of tools when designing interfaces, and to provide the necessary support to designers when they have questions.

6. Preview the built prototype

The "Prototype" referred to in this feature is slightly different from the common meaning of "Prototype". It suggests that there are some intuitive ways that designers can see how the Human–AI Interaction they create will work, such as how the flow of information in it will work. This part of the functionality will have a strong correlation with the UI level, Instance level, Message level and so on.

In addition to the 6 main features mentioned above, consider that this is a tool for designers who usually work in collaboration with others. Therefore there are the following additional features:

– Comment and annotate one or more messages
– Collaborative/Share/Export/Insert

These two features will help when groups of people are assisting to ensure that the team is on the same frequency, getting the same information, etc.

4.2 Ideas

Drawing upon the insights gathered from the testing phases discussed earlier, and taking into account the construction of design project functionality, it's time to explore the design process. This section will outline three specific concepts identified in the project, finally choosing one to drive the project forward.

4.2.1 Idea 1

The overall idea of the first concept is to "define each term according to the design use case, then use the term to build the Message and create the Human–AI Interaction". Therefore, there are two interfaces for defining terms using the specified content in the use case and for creating Message sequences using the terms, as shown in *Figure 24.1 & 24.2*. So here are two interfaces for defining terms using the content specified in the use case (*Figure 24.1*) and for creating Message sequences using the terms (*Figure 24.2*), respectively.

Figure 24.1. The Interface for defining terms.

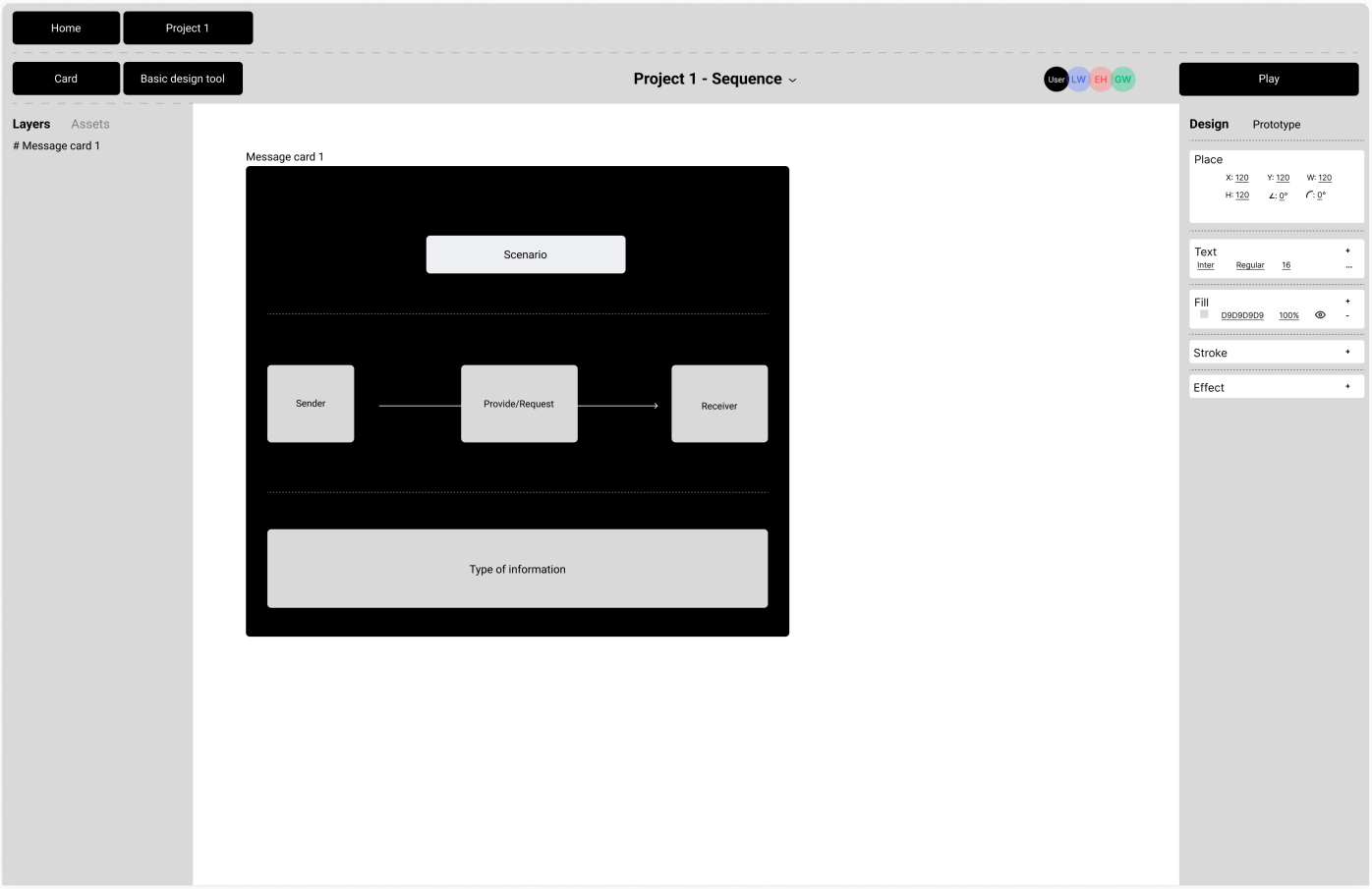


Figure 24.2. The Interface for creating Message & Message sequences.

The interface for defining terms is introduced first. Overall, there are three workspaces and one toolbar. See *Figure 24.3*. The workspace is divided into four independent areas to facilitate the term creation process in the current project:

Area 1 (left): This section provides an overview of various terms established in the project, including human roles. Designers can easily navigate, search, or check the status of term creation in this workspace.

Area 2 (center): Positioned in the center, this is the interface specific for the term creation. It offers an overal view of the terms that have been created.

Area 3 (right): Serving as the content modification area for different terms, this section enables designers to modify content for better alignment with the context. Each term has its own content modification area.

Area 4 (Toolbar): Located conveniently, this toolbar allows designers to create blank cards for different terms within Area 2 with just a click.

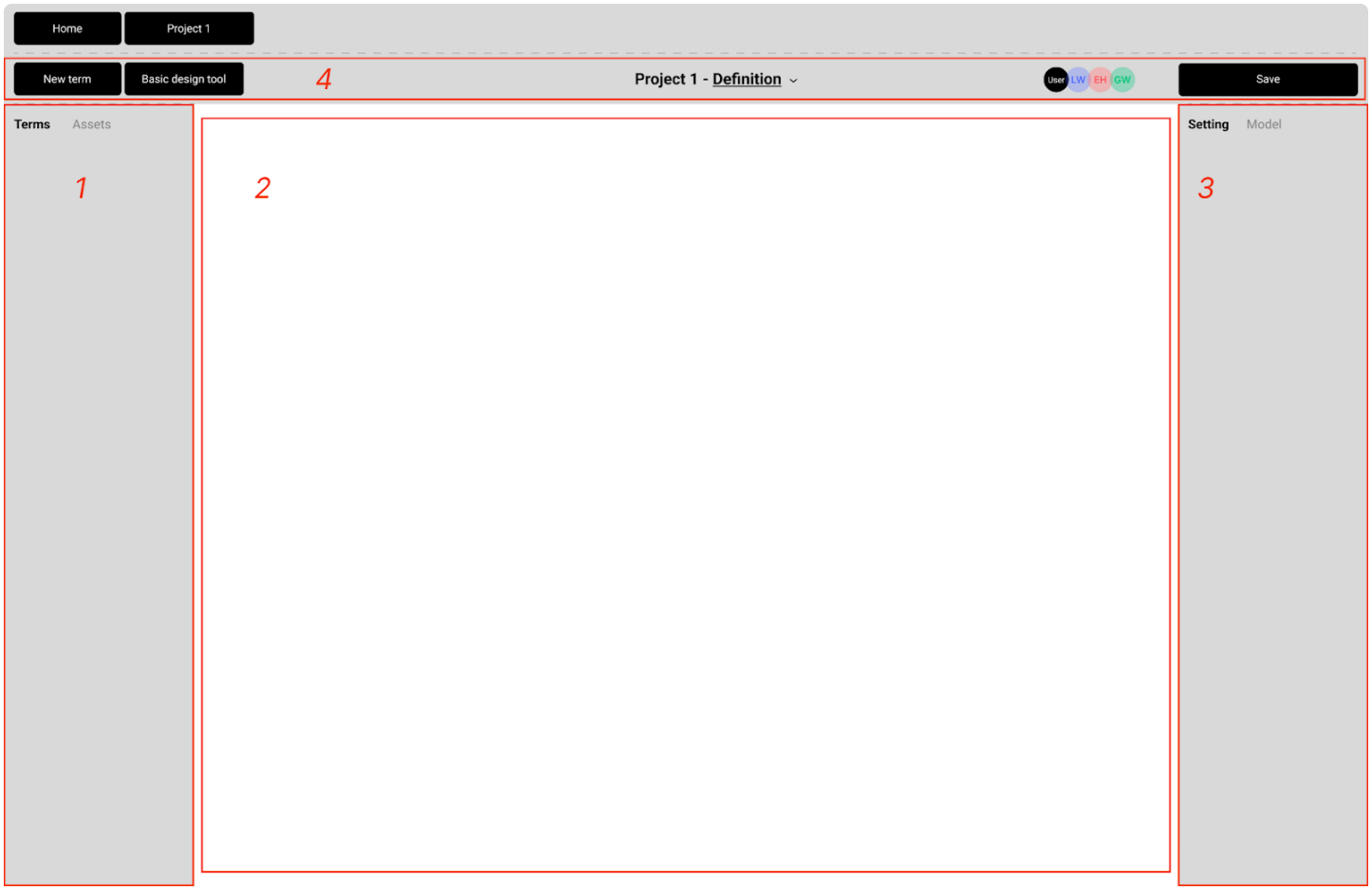


Figure 24.3. There are 4 parts in Definition Interface.

Depending on the context of the project, the terms that may need to be defined here are human role, model role, sample, two interpretations, feedback on human role or model role, etc. As shown in *Figure 24.1*. And there is one overview of the whole interfaces in *Figure 24.5*.

Then it is the introduction of creating messages and sequences. See *Figure 24.4*. Here again there are three workspaces and a toolbar. The workspace is divided into four distinct areas to facilitate the Human–AI Interaction creation process for the current project:

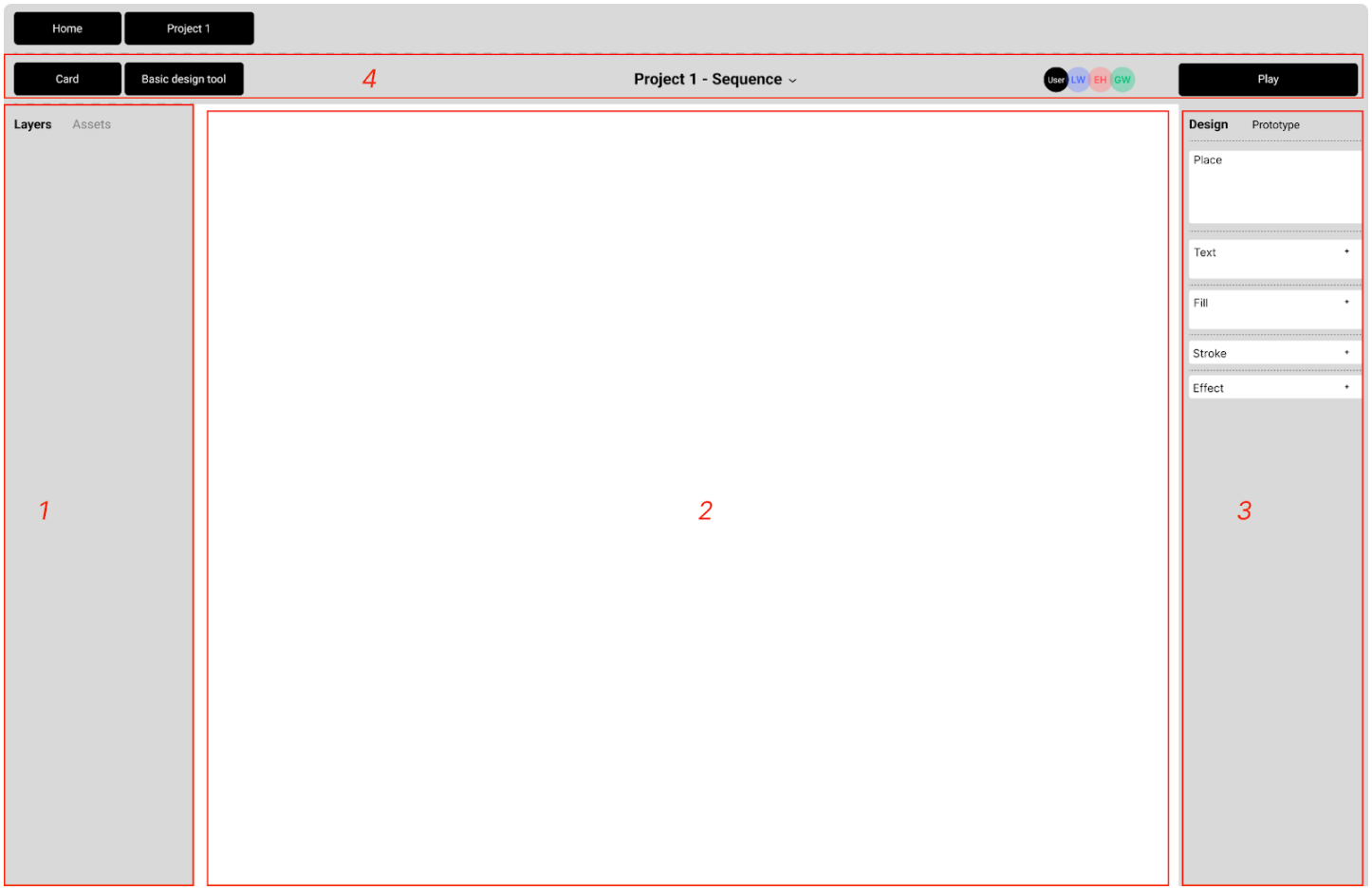


Figure 24.4. There are 4 parts in Sequence Interface.

Area 1 (left): This area gives an overview of the created messages in the project. designers can easily navigate, search or check the created messages in this workspace.

Area 2 (center): This interface is centrally located and is dedicated to presenting the created messages as well as the sequence of messages. Here it is possible to see the details of all messages and to change different message sequences or connections between messages.

Area 3 (right side): Here the designer can modify the details of the Message or create sequences or connections to other Messages.

Area 4 (Toolbar): This toolbar is conveniently located so that the designer can create a blank Message for Area 2 with just one click.

Figure 24.5. Examples of Idea 1's interfaces.

On this basis, an improved version was redesigned based on the constraints imposed by the project context on the degree of freedom of the different terms. As in *Figure 25*.

Figure 25. Overview of Idea 1's improved interfaces.

In this version (*Figure 25*), the primary concept aligns with the original design. However, there is a notable adjustment in the treatment of elements such as samples, the two interpretations, and feedback on the human role or model role. These components are now presented as non–negotiable aspects, resembling a database from which designers can only select, rather than having the freedom to create them.

4.2.2 Idea 2

The second concept shares the developed idea with the first but places a greater emphasis on Human–AI Interaction itself. In this concept, everything, except for Message and Message sequences, serves the purpose of these two components. Designers are prompted to add Message or Message sequences only when necessary. At the same time, in order to minimize the number of interruptions to the designer's thinking, this concept strengthens the work carried out in different interfaces in the previous concept into the only one interface. In short, there is now only one interface, and all the functionality as well as definitions of terms are present when a Message or Message sequence is required. See *Figure 26.1*.

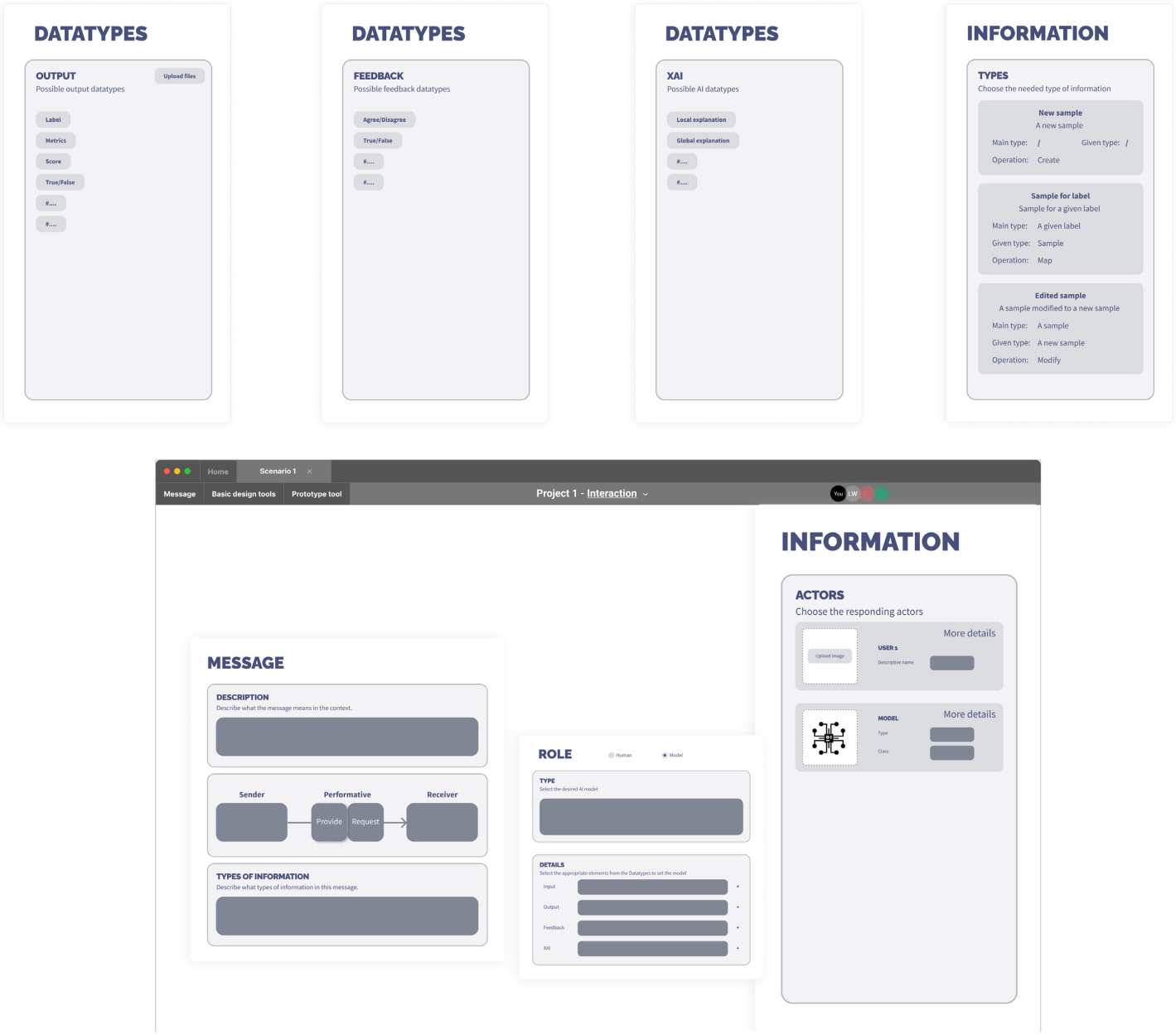


Figure 26.1. Overview of Idea 1's improved interfaces.

More details will be briefly described after some understanding of this concept as a whole. Here again there are 2 workspaces and a toolbar. See *Figure 26.2*. The workspace is divided into 3 different areas to facilitate the Human–AI Interaction creation process for the current project:

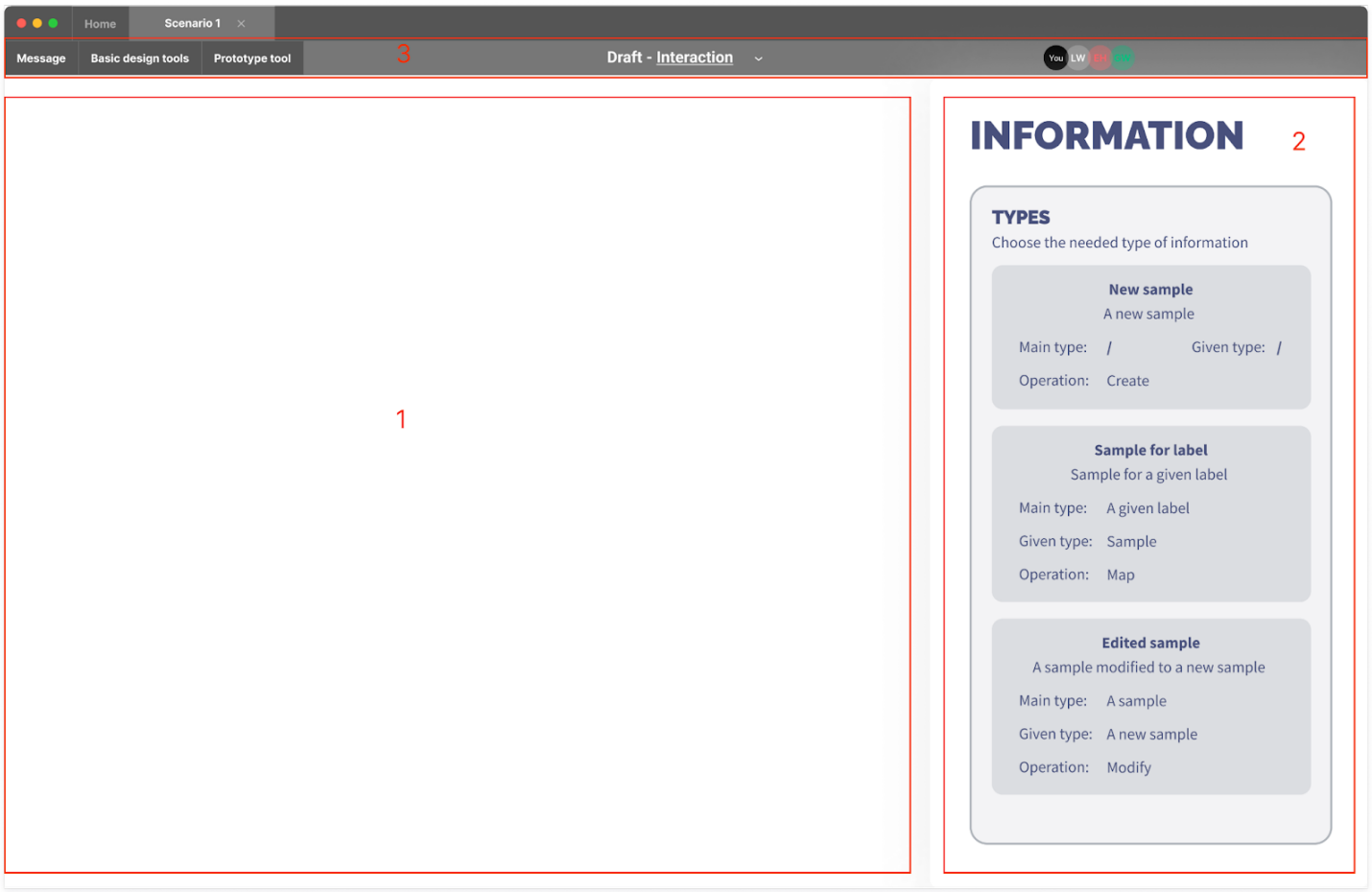


Figure 26.2. There are 3 parts in Idea 2.

Area 1 (center): This interface is located in the center and is used to display the created messages as well as the message sequences. Here it is possible to view the details of all messages and to change different message sequences or connections between messages.

Area 2 (right): Here the designer can modify the details of a message, or create sequences or connections to other messages. Different roles can be created here, including human or model roles, depending on the requirements of the information. See one example in *Figure 26.3.*

Area 3 (Toolbar): This toolbar is conveniently located so that the designer can create a blank message for Area 1 with a click.

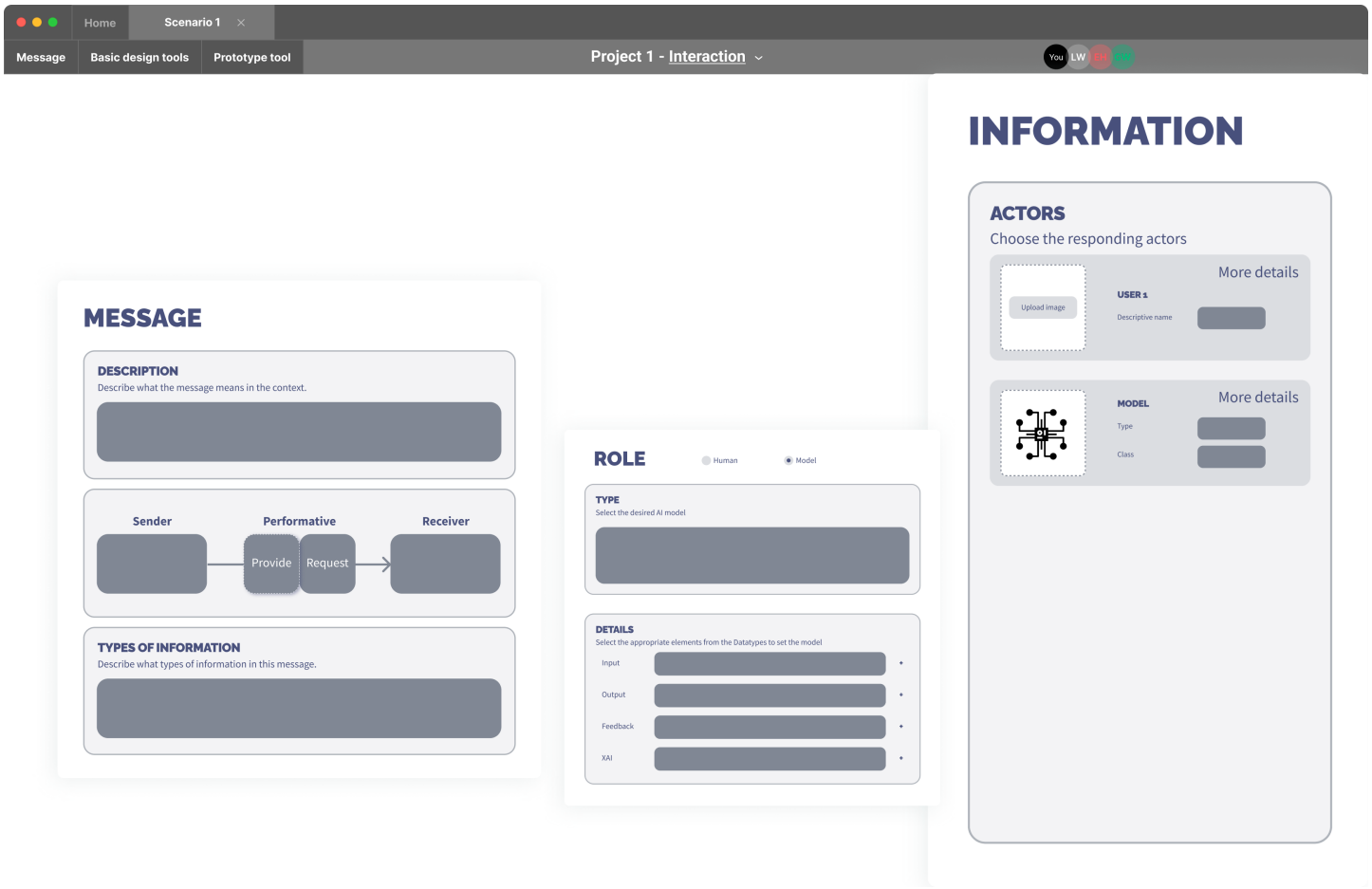


Figure 26.3. Creating a new role for Message.

4.2.3 Idea 3

Both of these types of ideas have the construction of Human–AI Interaction as their primary task in the underlying logic and do not provide designers with a view to design from the user's perspective. Considering the emphasis on "a Human–center design process" in the design goals and the user journey map, another common design tool for designers, inspired the next Idea 3.

While the user journey map is certainly not new to designers, a brief review of some of the user journey maps will make understanding Idea 3 easier before proceeding to its introduction.

The user journey mapping [38, 39] serves as an effective approach for comprehending relevant user processes, enabling the identification, and planning of essential user experience activities within a concise timeframe before delving into the user research phase. See *Figure 27.1*.

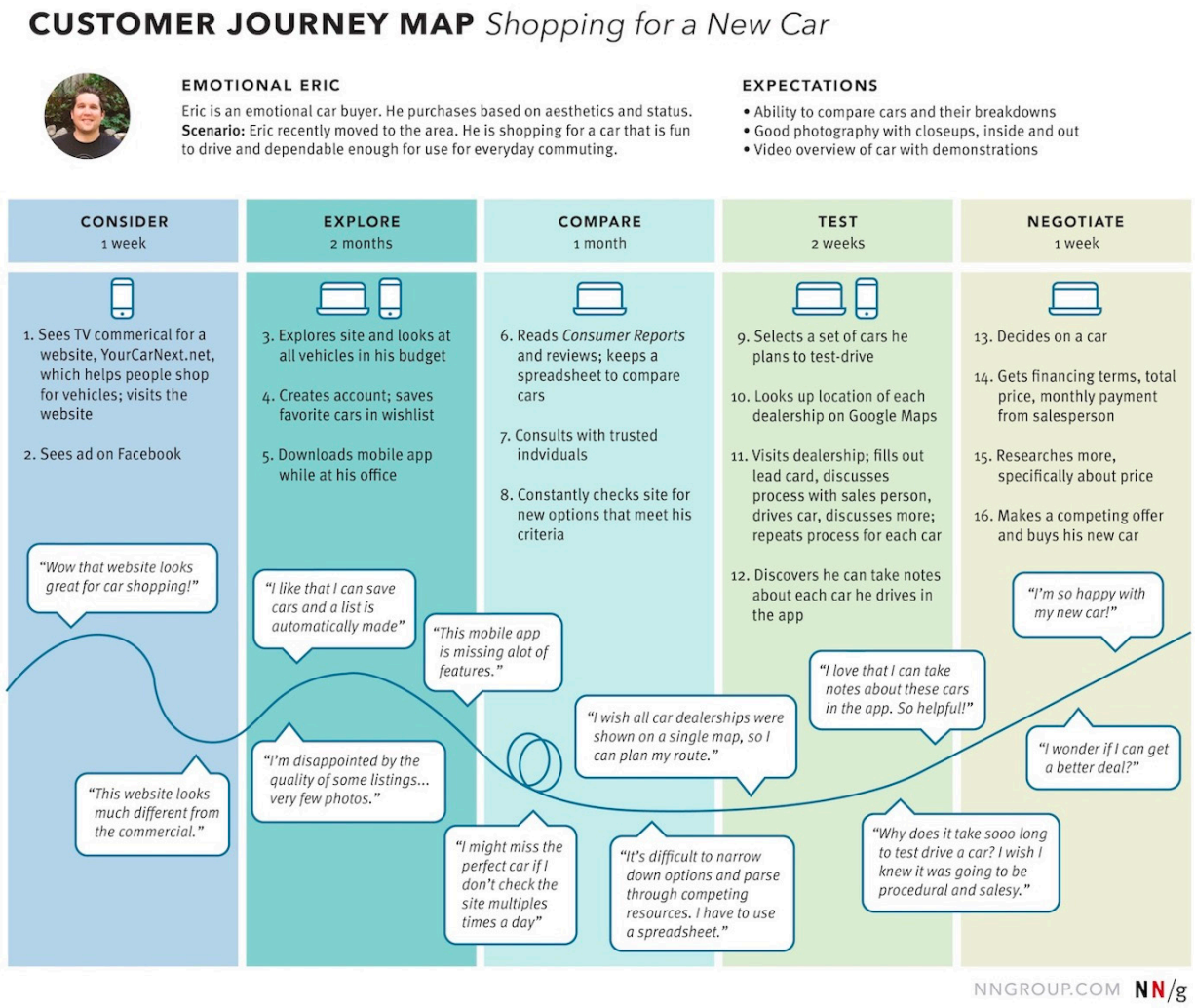


Figure 27.1. User journey map. From Nielsen Norman Group logoNielsen Norman Group [39]

In Idea 3, as shown in *Figure 27.2* below, the basic concepts in this project were mapped to the elements in the user journey map to create the interface as shown in *Figure 27.3*.

Concepts in Idea 3	Elements in the user journey map
Role	Target user's persona
Description & Message sequence	User step
Sender & Receiver & Actions	User actions
Type of information	Goals & experience
Other levels (Data/UI/Instance/...)	Touchpoint

Figure 27.2. Similarities between concepts of the project and elements of the user journey map.

In *Figure 27.3*, the first step emphasizes the need for an overview of the Scenario. Here, the Scenario refers to the context within which the designer intends to create Human–AI Interaction. Adopting a human–centered design perspective, this context typically encompasses the target users, stakeholders, and the interaction of the designated task with the environment or other individuals. Precisely defining and annotating this context aids designers in gaining a deeper understanding of the user, thereby promoting the creation of more effective interactions.

Concurrently, it's important to recognize that a design project extends beyond individual elements like a human, an interaction, or a context. Therefore, the categorization of various Human–AI Interactions in a project based on contextual distinctions proves more helpful for designers, enhancing efficiency in their work.

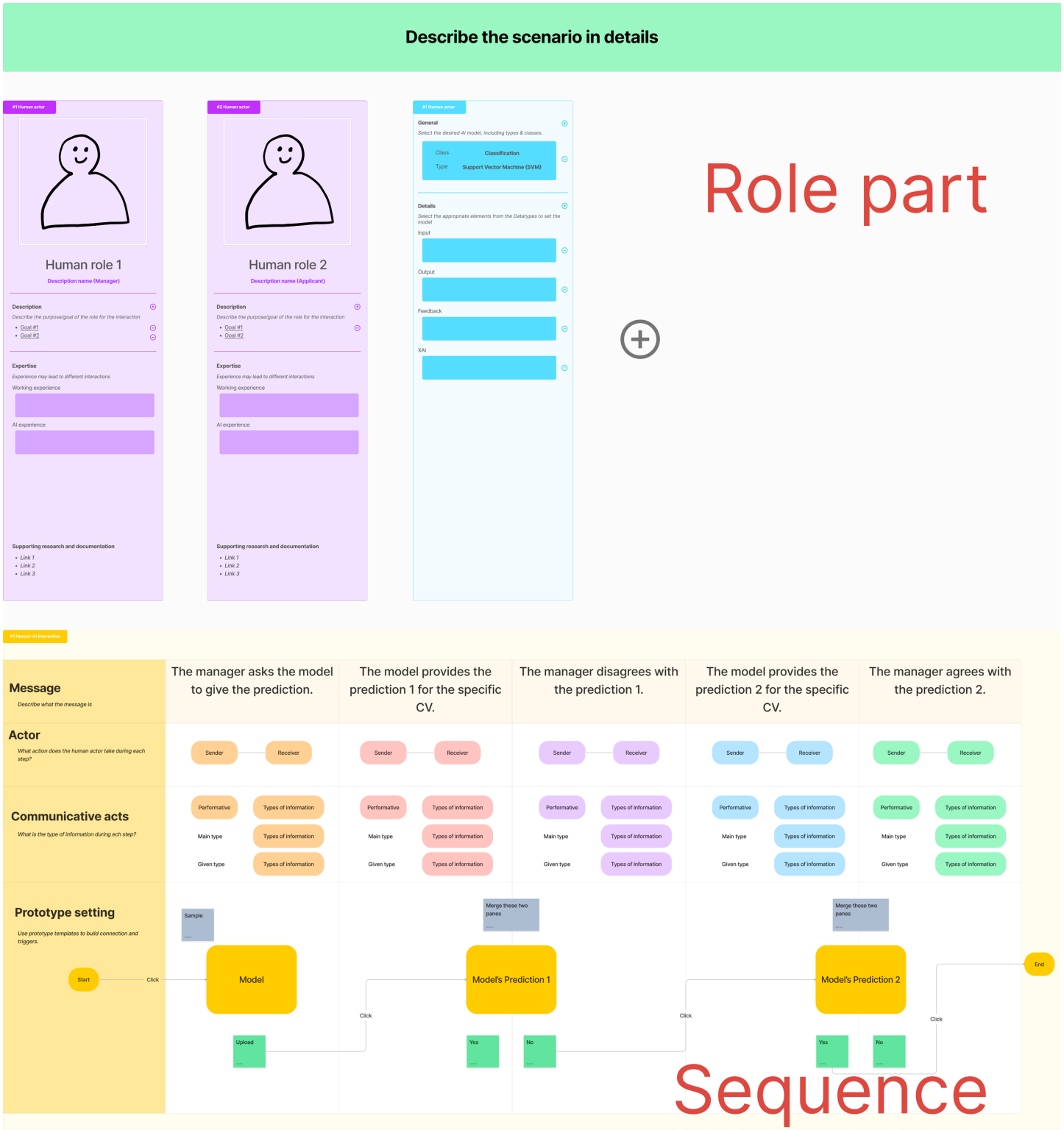


Figure 27.3. Idea 3. Inspired by Figma [40].

In *Figure 27.3*, there is a dedicated section for creating a Role part. Recognizing that "Humans" are important in the entire interaction process, establishing the possible roles becomes a step in the necessary preparations before designing. These roles include those mentioned above, like target groups and stakeholders.

This intentional setup assists designers in maintaining a focus on the human elements when crafting subsequent Messages and Message sequences. It establishes a foundational understanding of the key roles involved in the Human–AI Interaction, ensuring a human–center design approach throughout the design process, with reference to approaches such as Persona.

In *Figure 27.3*, the "Sequence" section adopts a tabular format, drawing inspiration from the user journey map to maintain a top–to–bottom hierarchy of information. This design choice is intentional and serves two primary purposes.

Firstly, presenting the information in a user journey map–like format helps bridge the unfamiliarity that designers may have with using Message sequences to represent Human–AI Interaction. By aligning this new concept with a familiar tool, the learning curve is likely to be more manageable for designers, minimizing the associated learning costs.

Secondly, based on the Information Architecture (see *Figure 23.1*), the "Description" serves as the annotation for the Message and should be one important featured at the beginning. This strategic placement aids designers in understanding other content by providing a contextual foundation for balancing the complexity of the Message sequences.

Moreover, adopting a table–like format addresses the challenge of organizing the Message sequences and aligns them with the natural flow of information in a sequence. Tables inherently convey a sense of information passing from the beginning to the end of the table, aligning with the sequential nature of message interactions. This visual style helps designers in conceptualizing the directionality of information flow within the sequence, enhancing their understanding of the prototype and developing a more intuitive understanding of the communications between humans and AI models. Also to more clearly indicate this sequentiality, each Message will still be identified sequentially.

4.2.4 Final idea

These three ideas did not evolve simultaneously. The first and second ideas underwent numerous iterations, as shown in *Figure 28*, before the emergence of the third idea. And here is one link for the clickable prototype: Idea 1 & 2

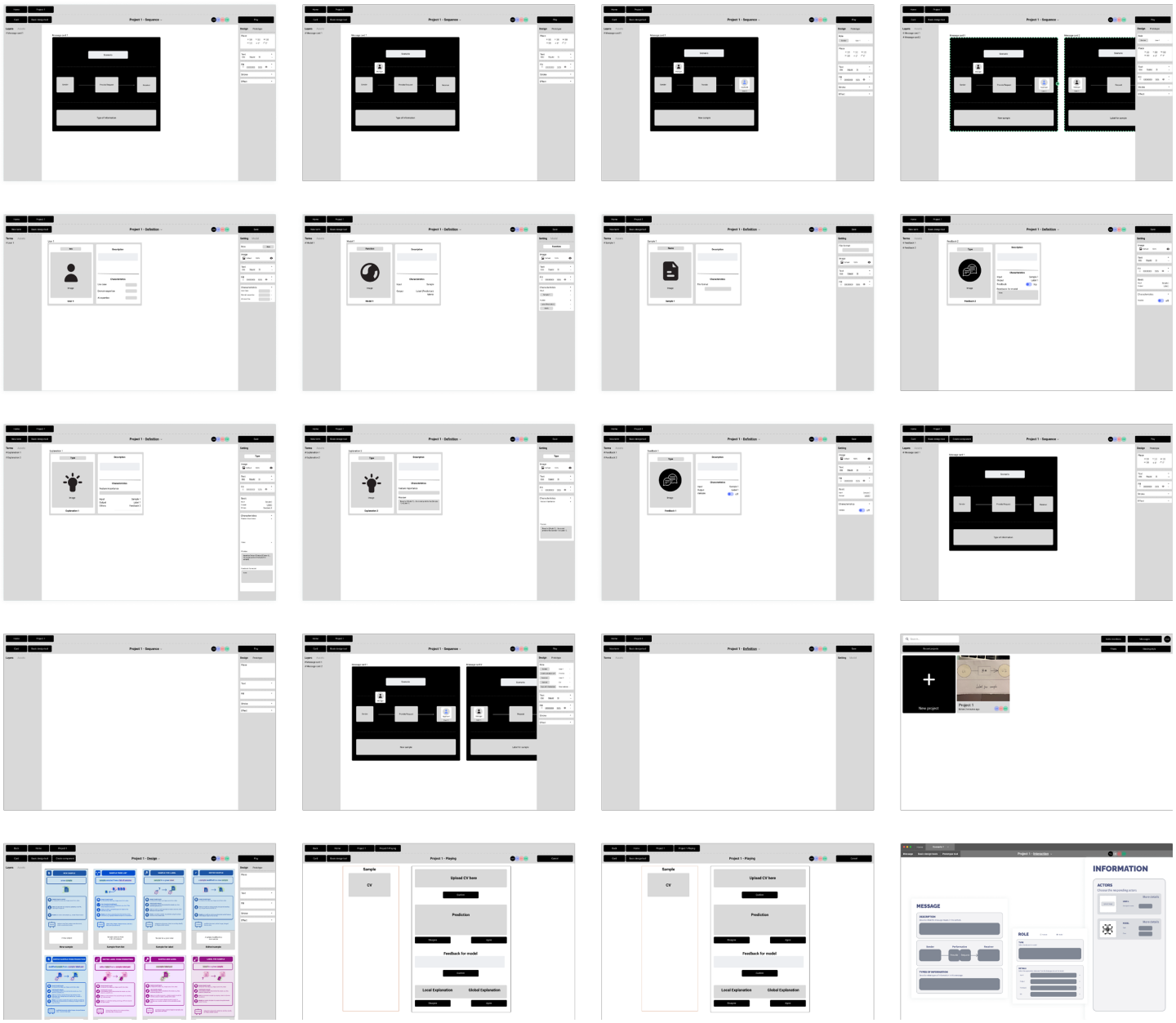


Figure 28. Overview of how Idea 1 & 2 went.

As shown in *Figure 29*, each had its own strengths and weaknesses. These four goals were chosen because they are relevant to the evaluation of design goals such as "Human-center design process," "Easy to learn how to use the tool," and "help. and "Help designers learn the basic concept “Communicative acts” of Human-AI Interaction involved in this project" in the evaluation of design goals.

While the first two ideas proved valuable in promoting designers' transition from specific design cases to the broader process of creating abstract interactions — an insight gleaned from the Pre-Phase & Phase 1 — they mainly focused on the Human-AI Interaction (HAI) itself and the flow of data within the interaction. Unfortunately, they overlooked the important human element. Moreover, both concepts introduced more specialized terms that required creation and definition by the designer.

In contrast, the Idea 3 aligns with the human-centered design goals by referring to the user journey map tool commonly used in human-centered design. It prioritizes human needs, introducing new concepts like communicative acts and functions from the design considerations. In a short word, it meets the design goals more than others.

As a result, the final design will be based on Idea 3.

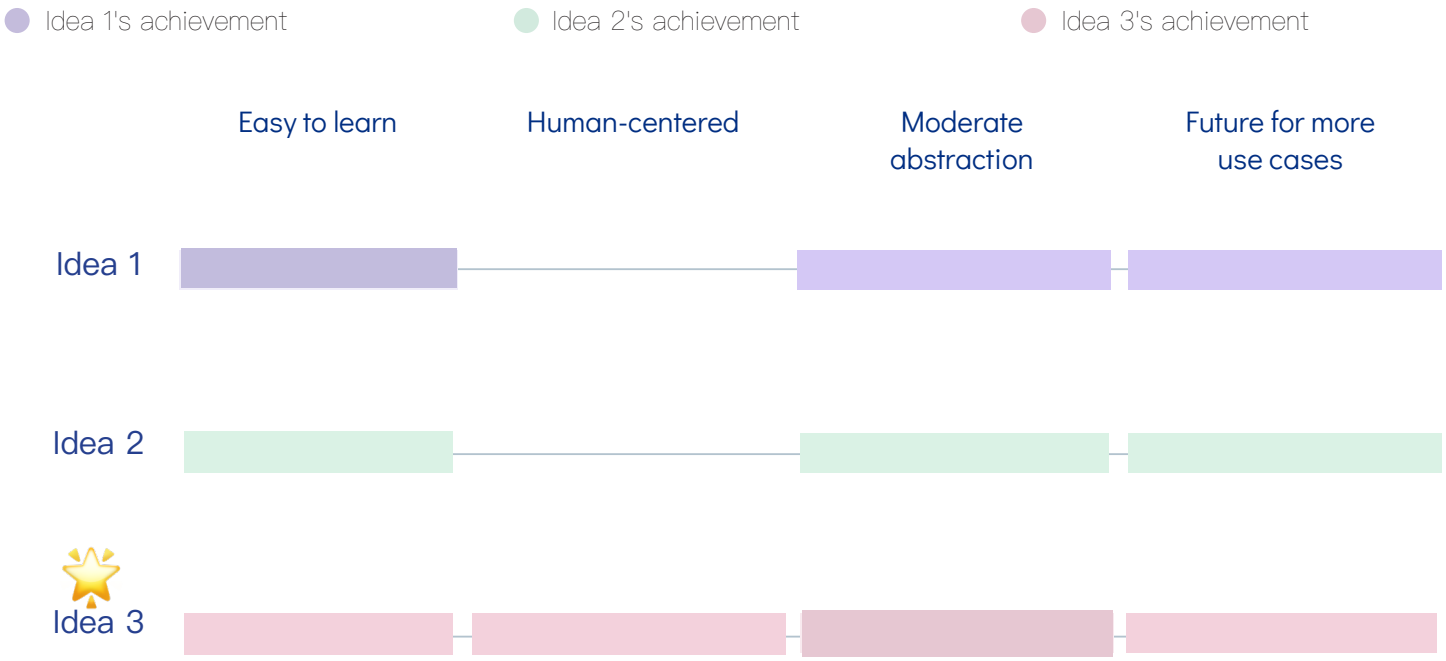


Figure 29. Overview of how Idea 1 & 2 went.

4.3 Conclusion

In this section, the design goals were refined, and the necessary functionality for the design outputs is specified. At the same time, all project–related terms were clearly defined, ensuring that researchers maintain a consistent level of understanding and could progress with the project cohesively. The Information Architecture between terms was established, laying the groundwork for subsequent design activities. Finally, among the three design ideas which met the 5 design considerations in Chapter 3, the one most aligned with the design goals is selected.

While the first two ideas held goods and poors extensive building and iteration, the decision to proceed without paying more attention to them during the design process was significant. This choice emphasized a forward–looking, goal–oriented approach, prioritizing options that best aligned with the project's goals.

5.
Test for evaluation

The fifth section will present a prototype that improves upon the previous section, Idea 3, and produce an interactive prototype for use in this phase of testing–Test 3. The exact process is shown in *Figure 30*.

At the same time, this section will only go into detail about the features that have been implemented for interaction; many of the designs so far have not been able to be put on the agenda for detailed discussion due to time limitation. In the future, there will be opportunities to gradually improve.

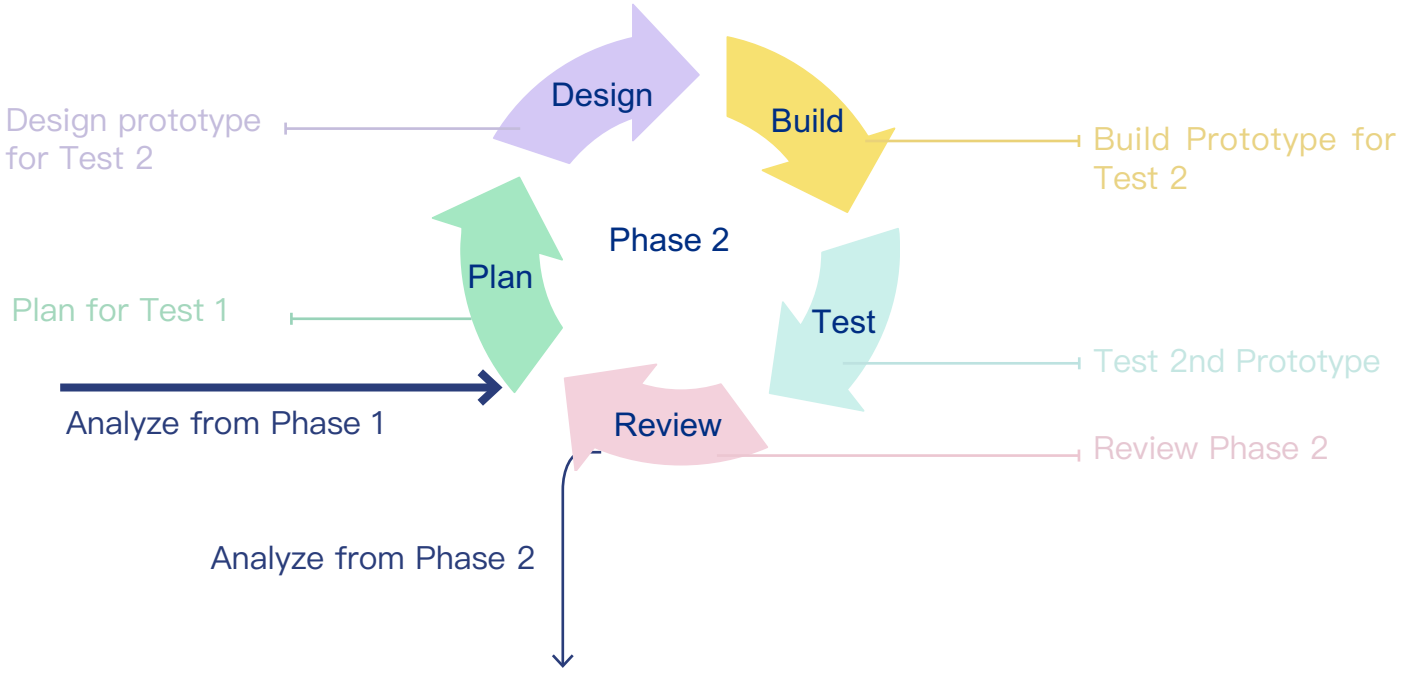


Figure 30. Overview of activities in Phase 2.

5.1 Overview

The prototype was optimized based on the concept of Idea 3, with a layout of different functions based on Human–centered design.

Figure 30.1 & 30.2 & 30.3 represents the idealized interfaces which hold all features. However, it's essential to note that many features shown in the figures have not yet been developed. At this stage, the main and completed functions consist of interactions related to Message/Message sequences, including creating, modifying, deleting, changing the order etc.

And some icons in all prototypes are from the Riddle Icon Pro library [41]. For the storyboards which were used in tests, visualized elements are from Freepik [42]. For the human figures in the poster used in tests, they are from Ayush Shakya in Figma [43]. For personas about the HR manager and Applicant 1 & 2, they were edited based on UXIS's work on Behance & Figma [44].

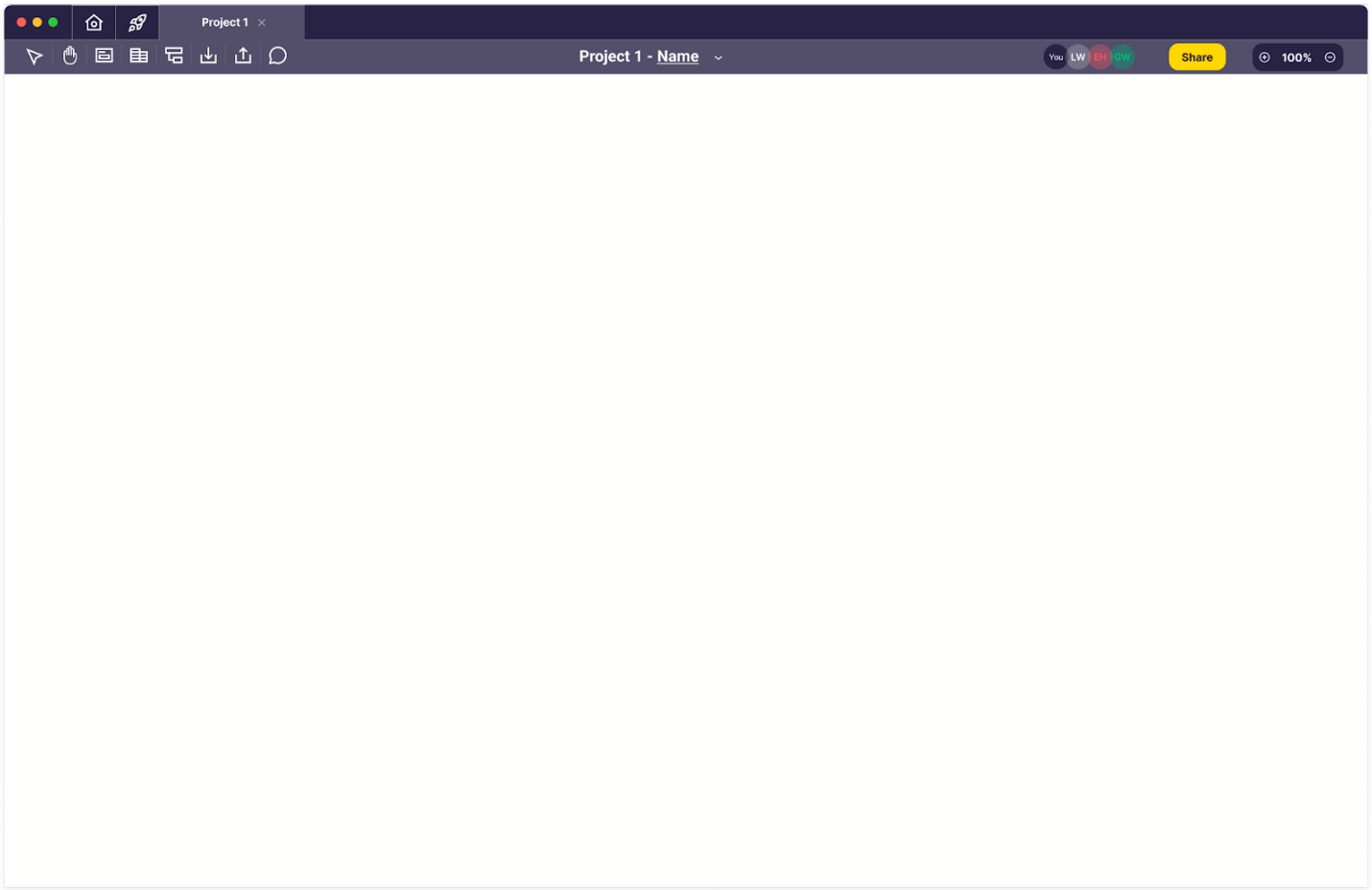


Figure 30.1. The envisioned completed interface – Blank.

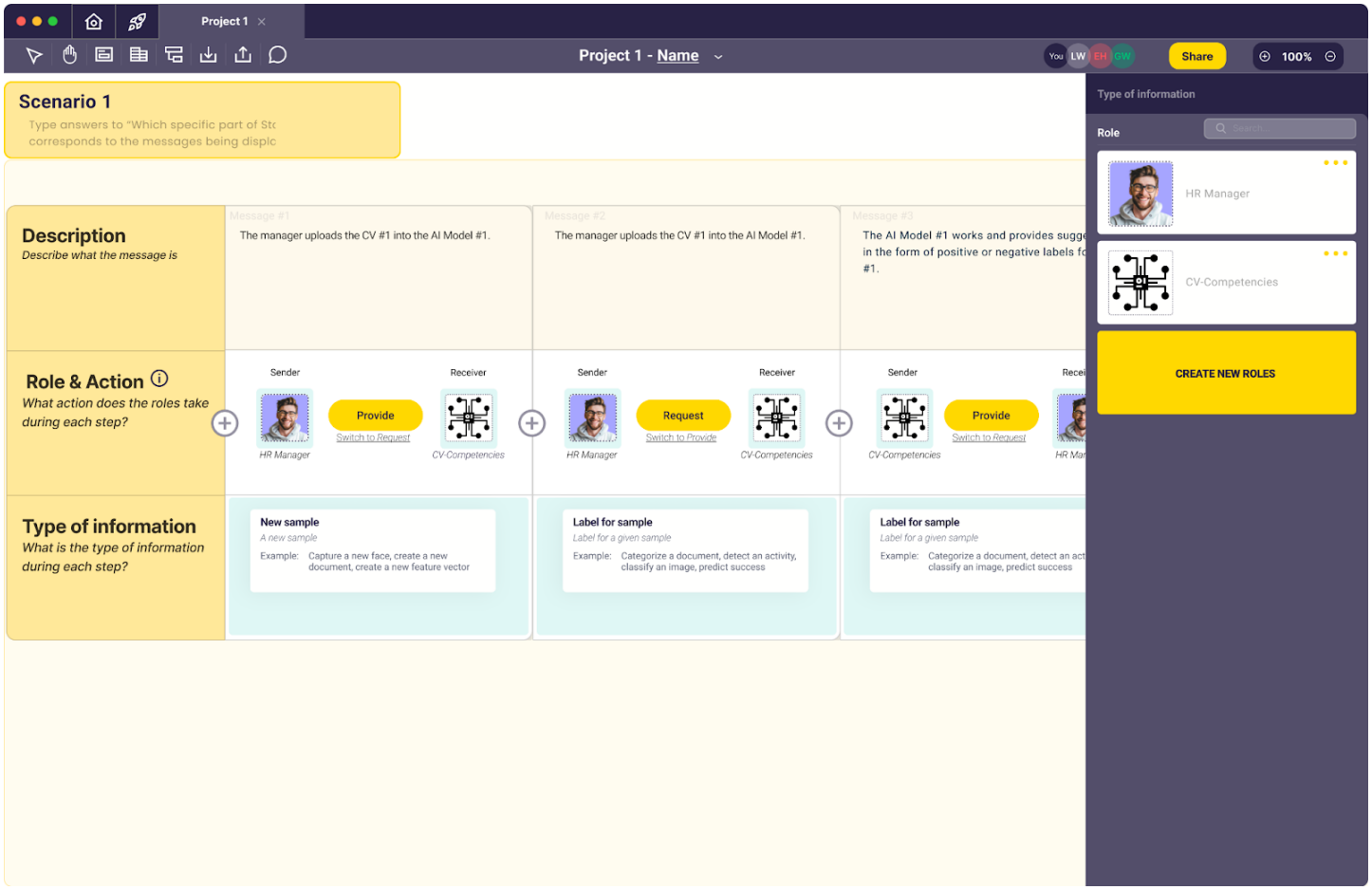


Figure 30.2. The envisioned completed interface – Message sequence.

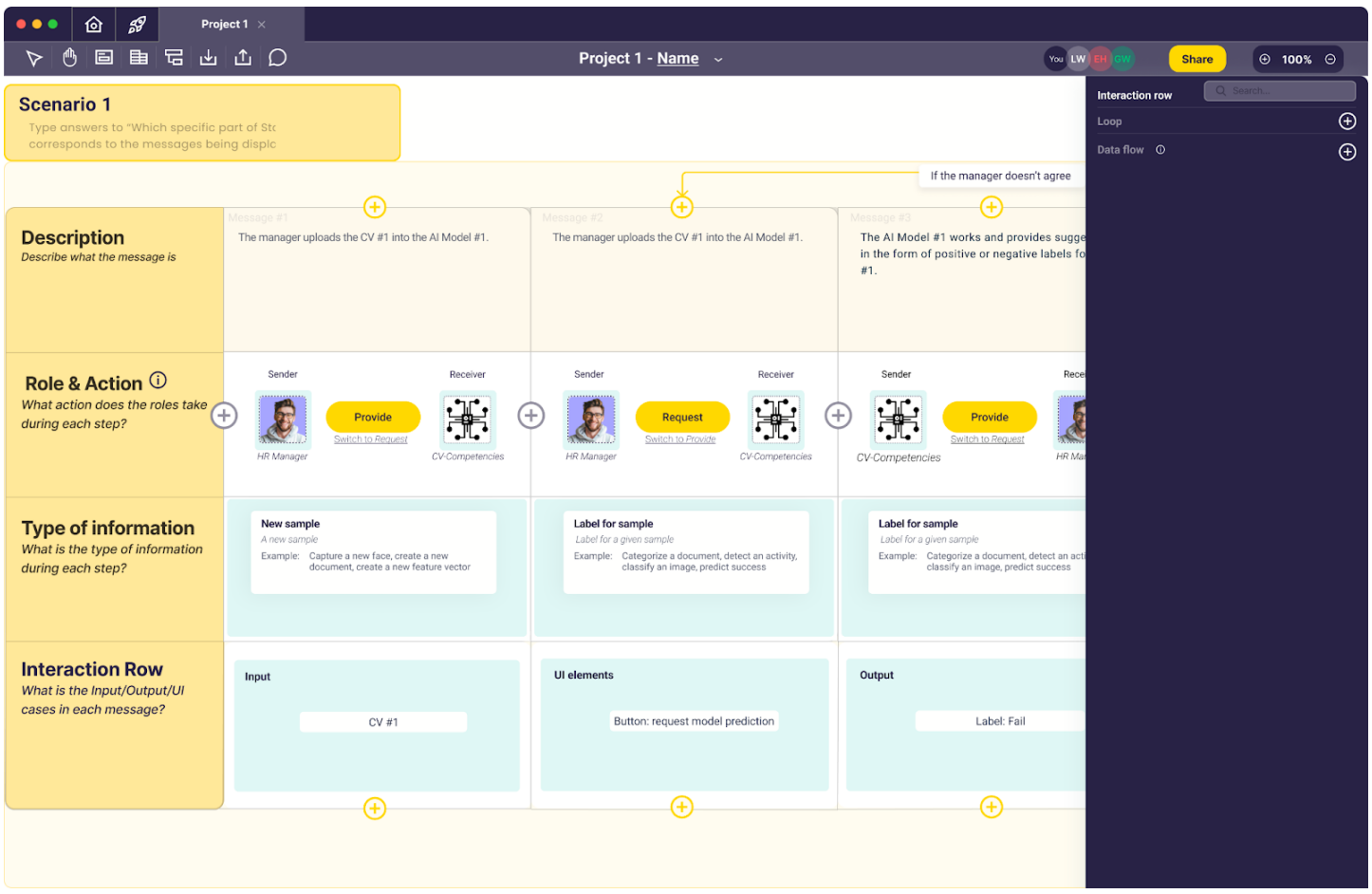


Figure 30.3. The envisioned completed interface – Showing interaction.

The preceding content provides a general explanation of the main interfaces, and this section briefly outlines another part of the interface dedicated to the user's personal workspace post-login. It's important to note that this section isn't the primary focus of the design, so the current version serves as a reference only. See *Figure 30.4*.

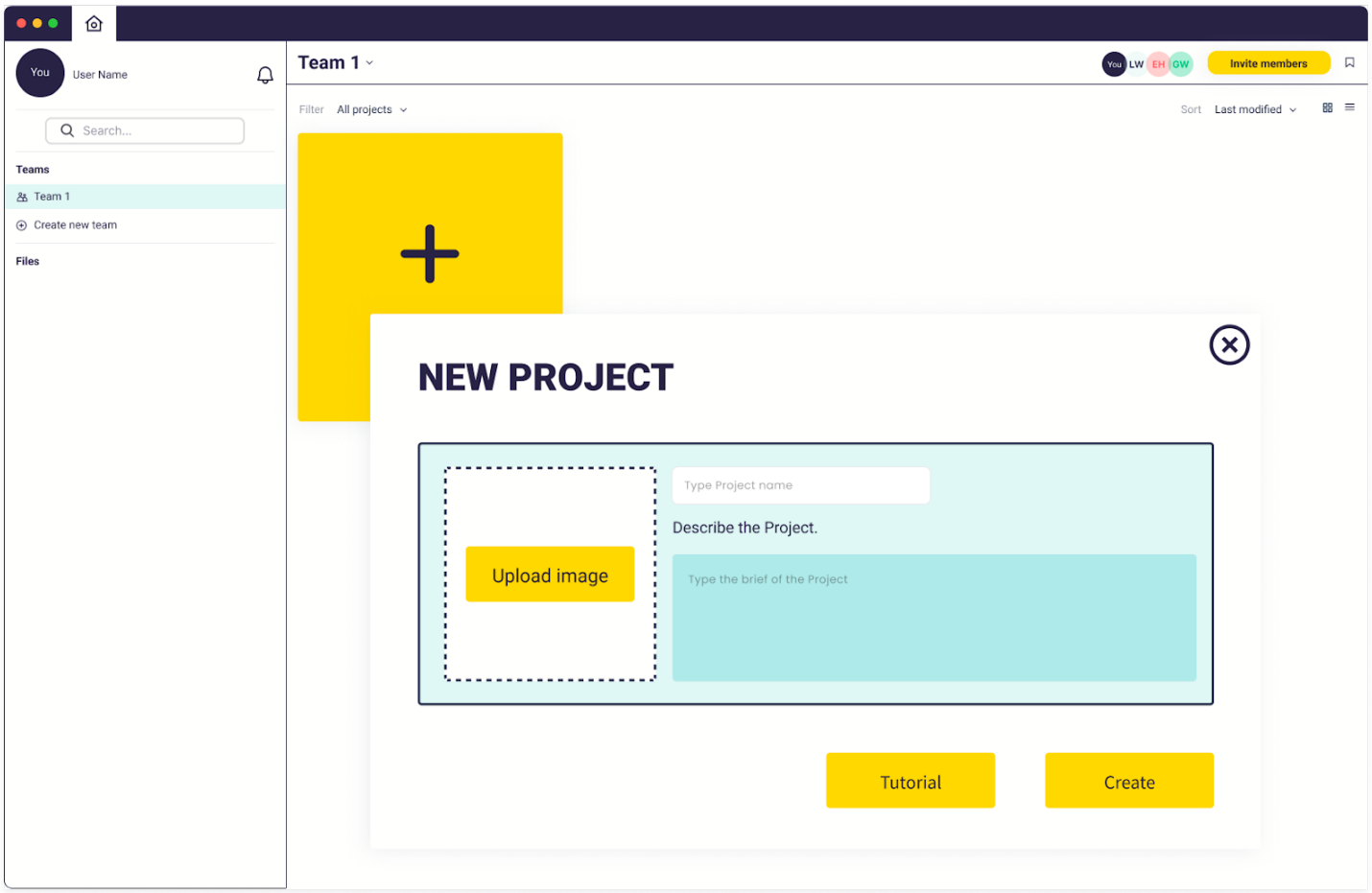


Figure 30.4. The workspace.

5.1.1 User interfaces

1. Top positioning status bar

Summarily, the user interface is consist of four main parts. Starting from the top, the first component is the Top Positioning Status Bar, shown in *Figure 31.1*. This bar serves to indicate the specific project the designer is currently accessing and offers a convenient means to navigate back to the tool's home page. Additionally, it may incorporate future development related to design inspiration.

The important poiny of this status bar lies in its functionality, enabling designers to promptly identify their current project and seamlessly switch between interfaces as needed.

Figure 31.1. The top bar

2. Toolbar

On the left side of the second level (shown in *Figure 31.2*) is the Tool bar.This is where almost all of the important functions of this interface are centralized. When using this bar, the following can be accomplished in order from left to right :

– Mouse for Move function
– Mouse for Hand tool
– Create Scenario
– Create Message (sequence)
– Switch to Connection mode and display content related to Input/Output/UI elements
– Importing files
– Exporting files
– Add comments or annotations

Figure 31.2. The toolbar.

3. Project name

In the middle of the second layer (shown in *Figure 31.3*) is the Project name. this locates the scope and content of the current interface work. Designers can change the project name and may have more features here in the future.

Figure 31.3. The project name.

4. Collaborative boards

The second layer, on the right, is related to collaboration and will show the collaborators who are working on this interface as well as buttons to share and invite more collaborators to join the project. See *Figure 31.4*.

Figure 31.4. The collaborative boards.

5. Scale display

Formed on the far right of the second level is a common tool: showing the percentage the user is zoomed in. It is not currently possible to interact with it, but zooming the canvas via gestures is a possible future implementation. See *Figure 31.5*.

Figure 31.5. The collaborative boards.

6. Bench

This is the area that will be displayed after the Scenario and Message (sequence) are created. This is an infinite canvas and can be freely moved or zoomed in and out. See *Figure 31.6*.

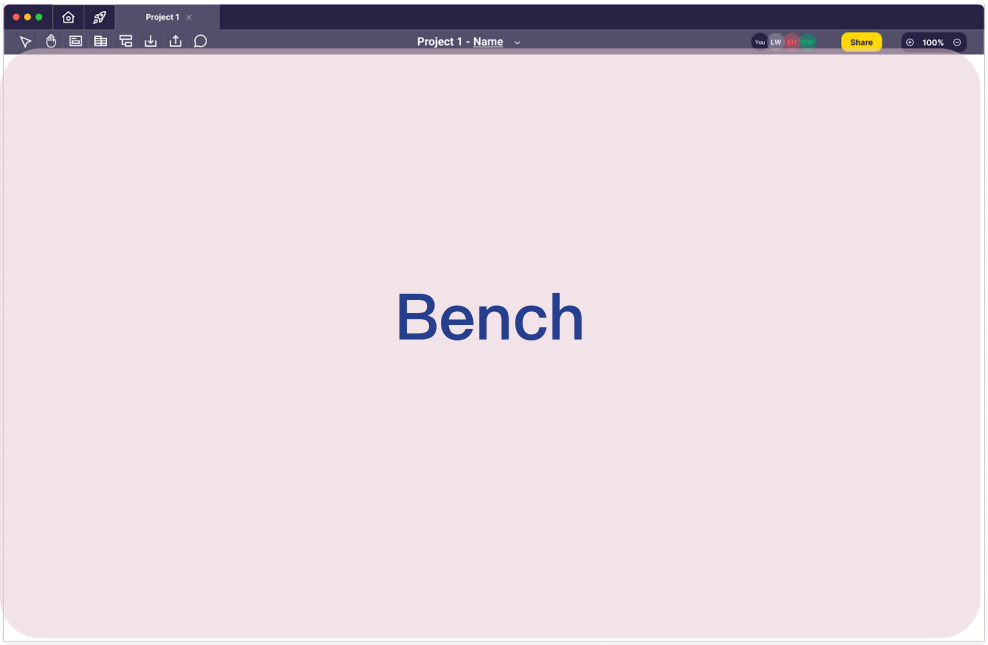


Figure 31.6. The blank bench.

7. Right function part

As shown in *Figures 30.1, 30.2*, and *30.3*, the dark-colored function part is always fixed to the right when there are Scenarios and Messages, and defaults to the Roles view (*Figure 30.2*), which displays the existing roles. When entering the "Interaction" version, the right function part defaults as shown in *Figure 30.3*.

More explanation on how this part changes would be shown in 5.1.2 General interactions.

5.1.2 General interactions

The tool prioritizes user-friendly and straightforward interaction methods, especially for users new to the system. To keep a balance, familiar and commonly used interaction methods have been combined. These include:

- Utilizing the mouse wheel for vertical scrolling.
- Enabling drag-and-drop functionality in all directions through the "Hand tool."
- Zooming in and out using the keyboard "Shift" in conjunction with the mouse wheel.
- Selecting content with the left mouse button, and dragging and dropping by long-pressing the left button.
- Summoning the menu with the right mouse button, offering options such as delete, paste, copy, etc.
- Alternatively, users can directly use the keyboard "Delete" key for deletion.

The prototype, used for the later test in this phase, only implements left and right movement functionality, as well as functions such as a left mouse click to select content or a long press to drag and drop content. However, there are plans to refine the above settings in the future.

5.2 General tools

While the previous section focused on explaining the interface and its working areas from a holistic point of view, this section will focus on introducing the main tools, including Scenario/Message (sequence)/Roles/Types of information/ Interaction, and so on.

5.2.1 Scenario

After creating a Scenario, here is one element shown in *Figure 32.1*. Scenarios serve the purpose of distinguishing between various Human–AI Interactions created by designers within a single project. The goal is to provide a contextual differentiation, offering an overview of each interaction to enhance designers' understanding of the interactions' significance. For instance, in a CV screening use case, the outcomes of an interaction designed for an HR manager collaborating with AI models might differ from those of an interaction designing for an applicant interacting with AI models.

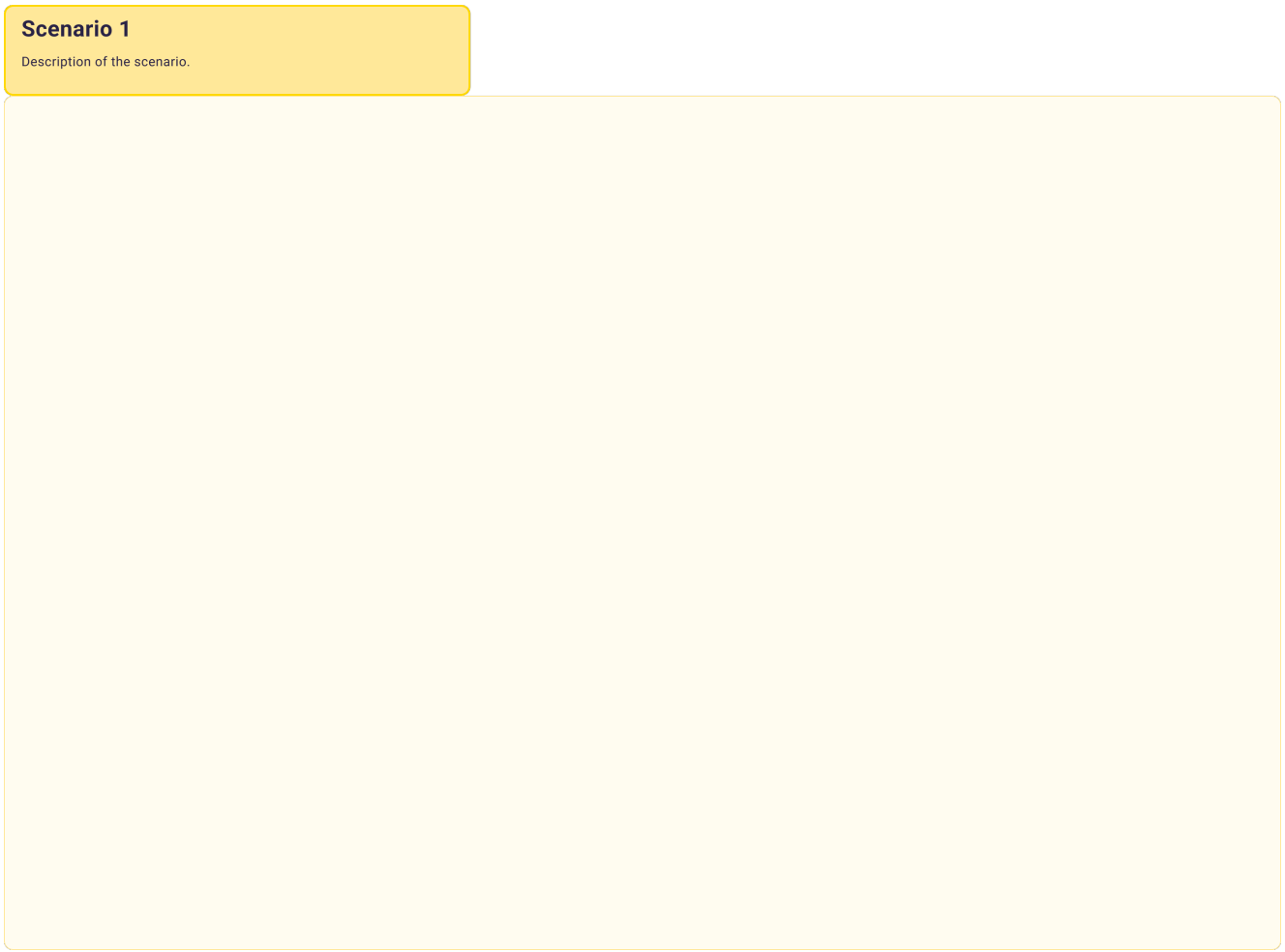


Figure 32.1. Scenario.

5.2.2 Message

A Message (sequence) is typically created within a Scenario. *Figure 32.2* illustrates an empty Message (sequence), while *Figure 32.3* showcases the Message (sequence) after being populated. To incorporate a new blank Message, utilize the "Add" button highlighted in the red circle in *Figure 32.3*.

As shown in *Figure 32.2 & 32.3*, there are three elements to complete in this section. The first row is designed for providing the meaning of what the Message is in this context, usually typied in natural language or something humans can read. This supports designers in readily understanding the meaning of the Message they own or others have generated.

The second row is filled with Sender/Receiver and the roles in between. The Sender means the role initiating the entire behavior or message, which could be a model or a human. The Receiver means the role of receiving the behavior or message, which may be a model or a human.

The yellow button between the Sender and the Receiver is for entering the behavior emitted by the Sender, with options for Action which contains Provide or Request.

The final row is filled with what kind of information exchanged during the communication. There are 12 types of information.

So how should designers fill in the Message? It will be explained in Roles and Types of information.

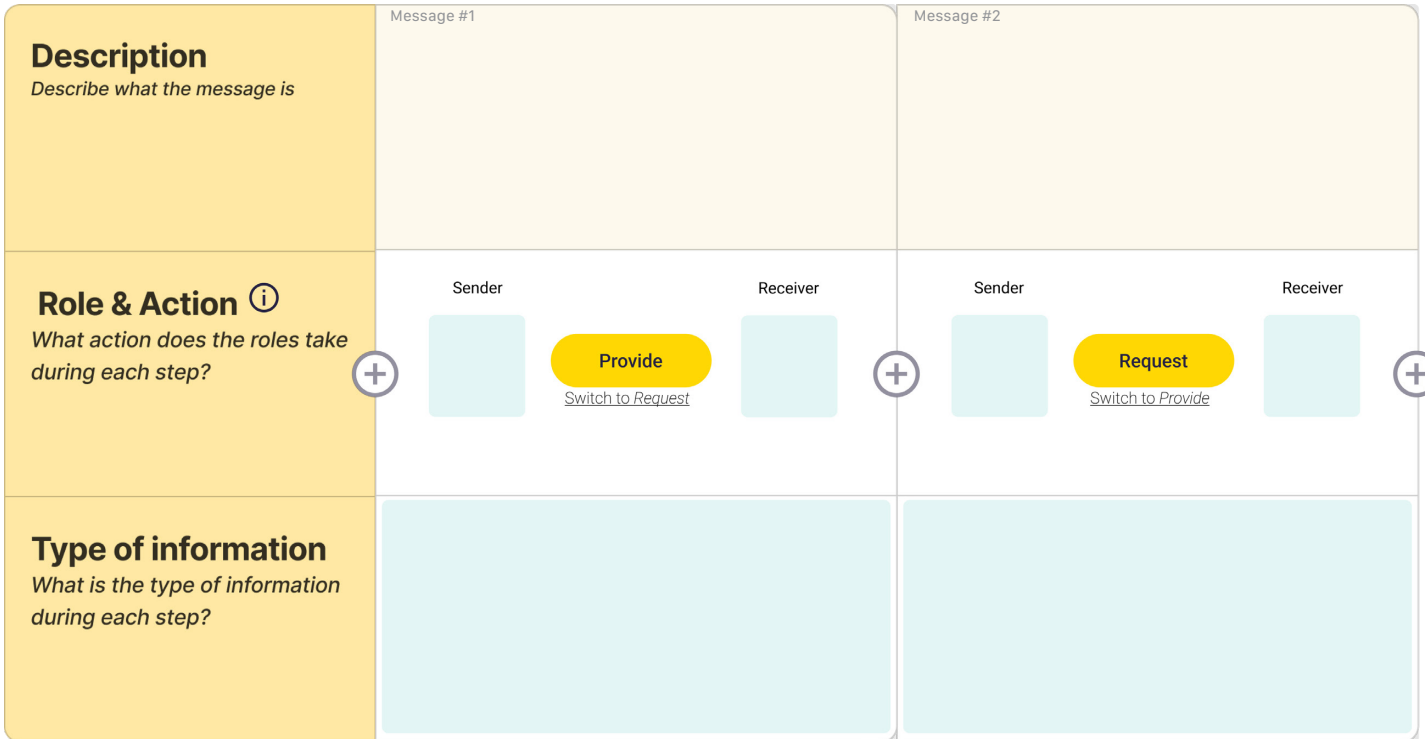


Figure 32.2. Blank Message (sequence).

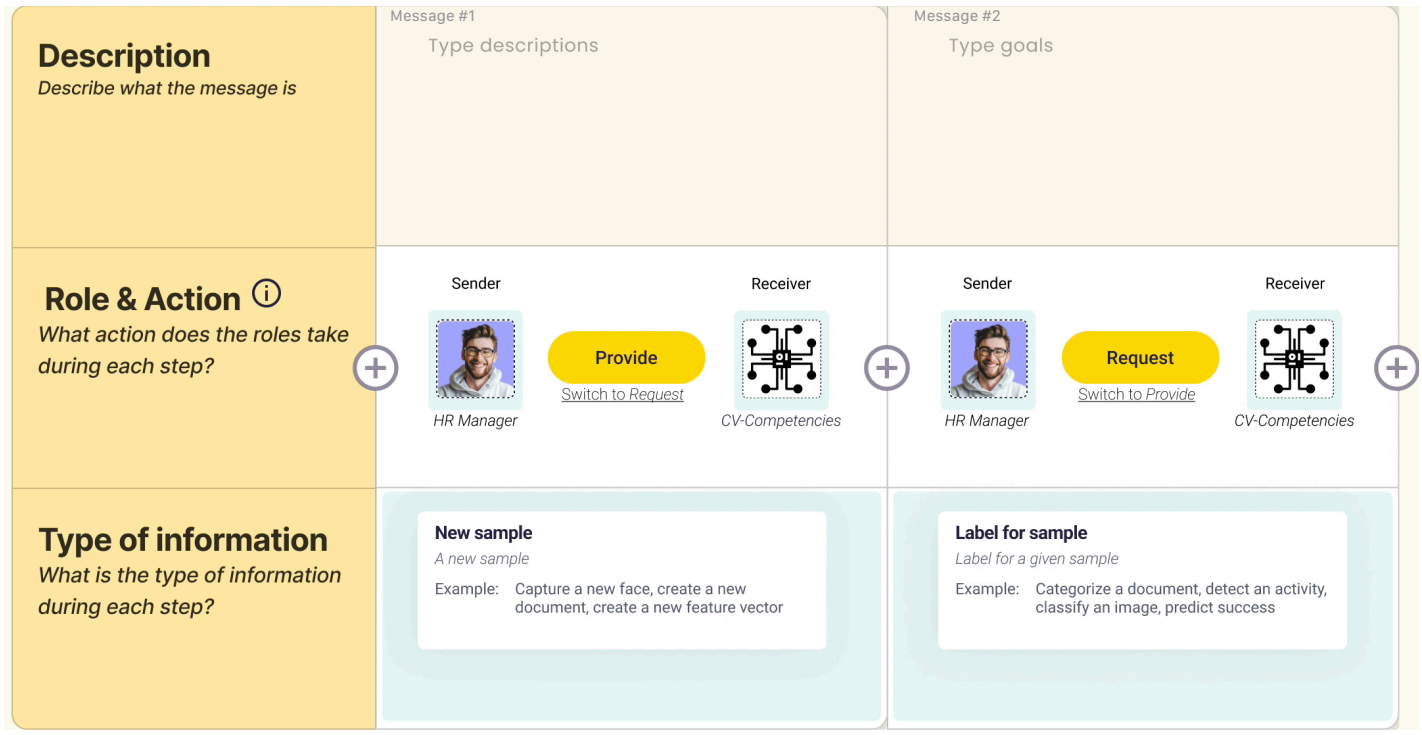


Figure 32.3. Filled Message (sequence).

5.2.3 Roles

1. Human Roles

For Human Roles, there are two different interfaces that need to be introduced here.

First, designers need to select or create a role from the right side function part of the interface, such as the content shown in *Figure 32.4*, and the overall interface is shown in *Figure 32.5*. Drag and drop the desired role in *Figure 32.4* to the section corresponding to Message (sequence).

Figure 32.4. Function part for roles.

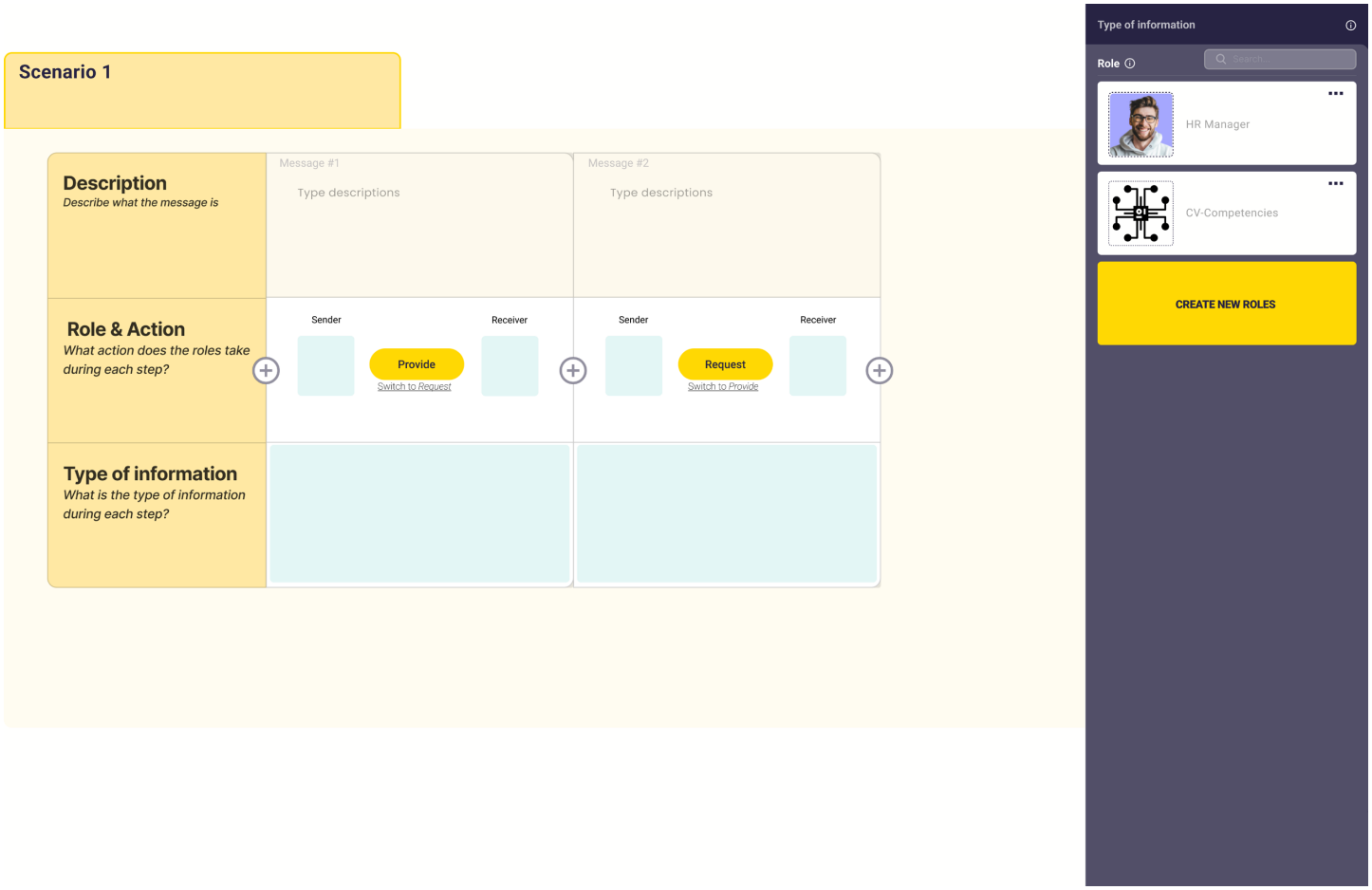


Figure 32.5. Whole interface for filling roles.

The second tool related to Human roles is the creation of needed Human roles. See *Figure 32.6*. In constructing the elements necessary to create a new persona, consideration was given to the Persona tools commonly utilized by designers during the design process, along with the Information Architecture discussed in the previous section. Finally, it was decided to fill in four sections here:

– the persona's name
– an overview of the persona's goals
– other fundamental information.

This aids designers in visualizing the behavioral characteristics of the target user, fostering empathy, and assisting in the determination of user needs.

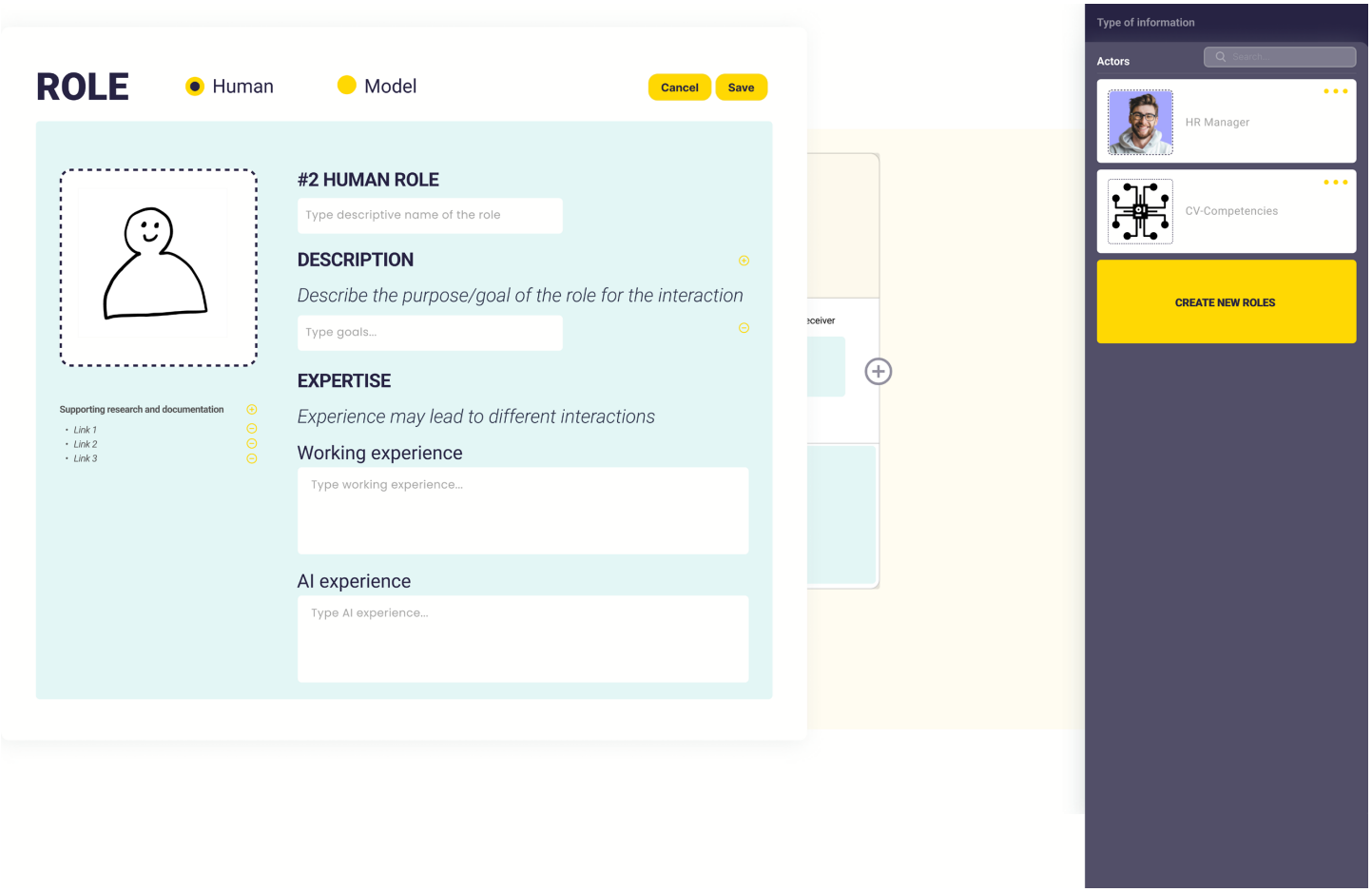


Figure 32.6. Creation human roles.

2. Model Roles

For Model Roles, two different interfaces need to be introduced here. One is when the designer needs to fill in the Role & Action in the Message (Sequence), this interface is the same as Human Roles (*Figure 32.4*).

The second tool associated with Model Roles is the creation of the required Model Roles. Consider the insights gained from tests, as well as the Information Architecture discussed in the previous section, when building the elements needed to create a new role. It was finally decided to fill in three aspects here: the persona name, an overview of the persona's goals, and a selection of pre-defined functional models. This reduces the cost of learning about the different model features for designers, who only needs to select them according to his or her needs, as shown in *Figure 32.7*.

So, how should a designer choose a preset function AI module? While filling in the Model Roles content, the overall interface is shown in *Figure 32.8*. The right side will display the different function models. Simply drag and drop the desired model to the corresponding position through Drag & Drop interaction. A model can only have one function module now.

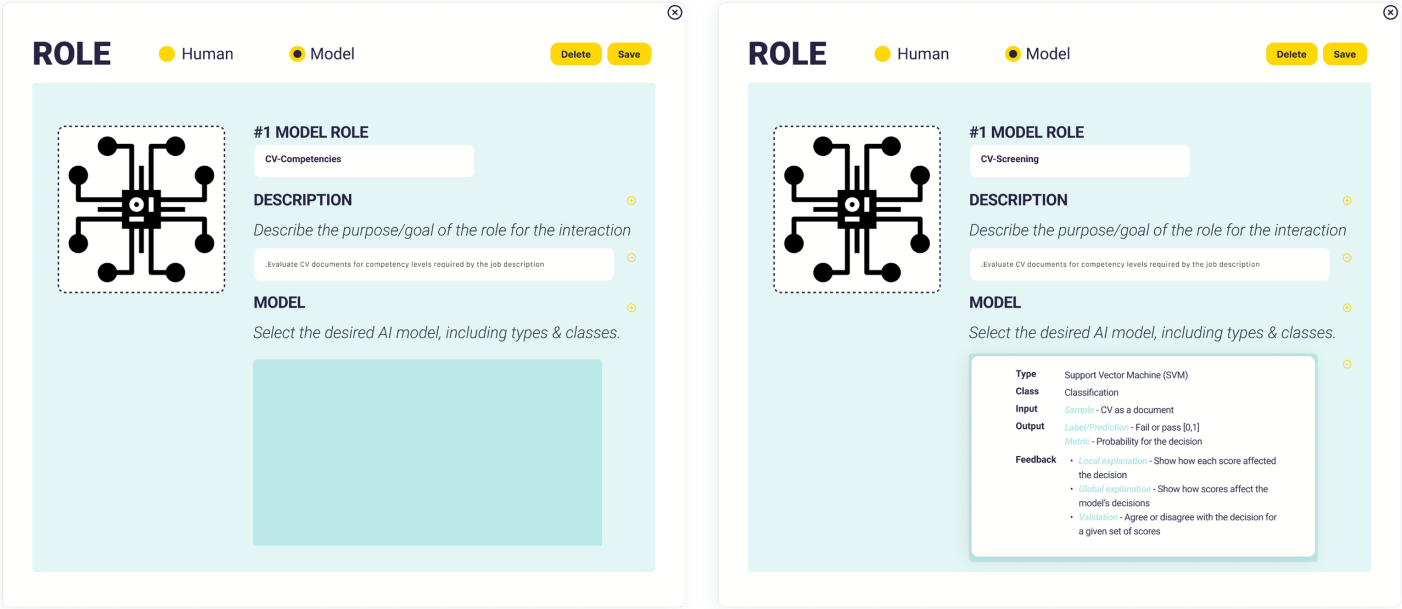


Figure 32.7. Creation Model roles.

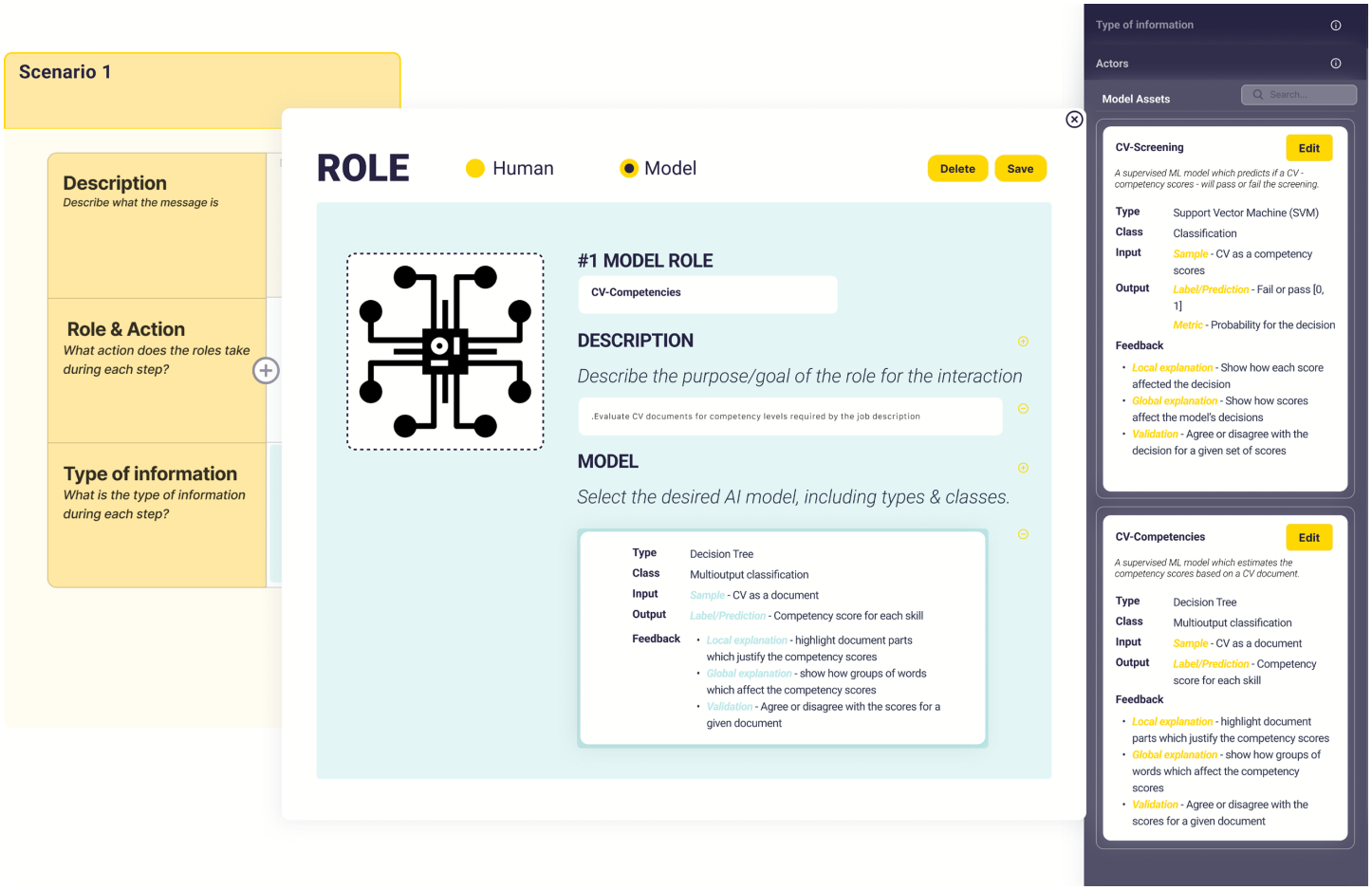


Figure 32.8. Choose function module for Model roles.

5.2.4 Types of information

When filling out the Type of information, the interface is shown in *Figure 33.1*. On the right, an overview list of 12 types of information will be presented, along with their names, meanings, and common examples. This may help designers quickly search and understand the meaning of each type of information, enabling them to select the one they need. At the same time, designers can determine what they want by reading detailed information about each type of information. See *Figure 33.2*.

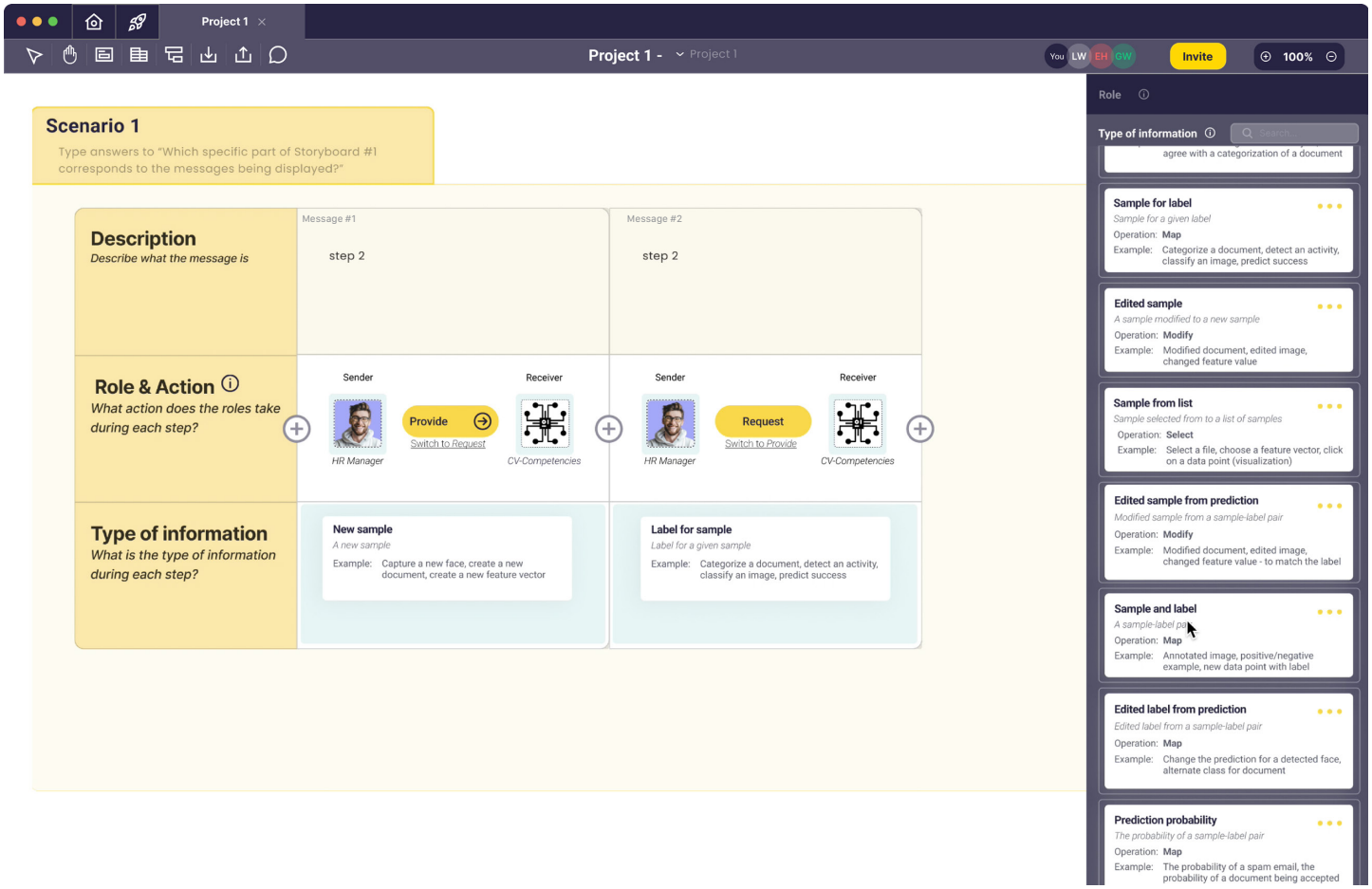


Figure 33.1. Fill Type of information from the right function part.

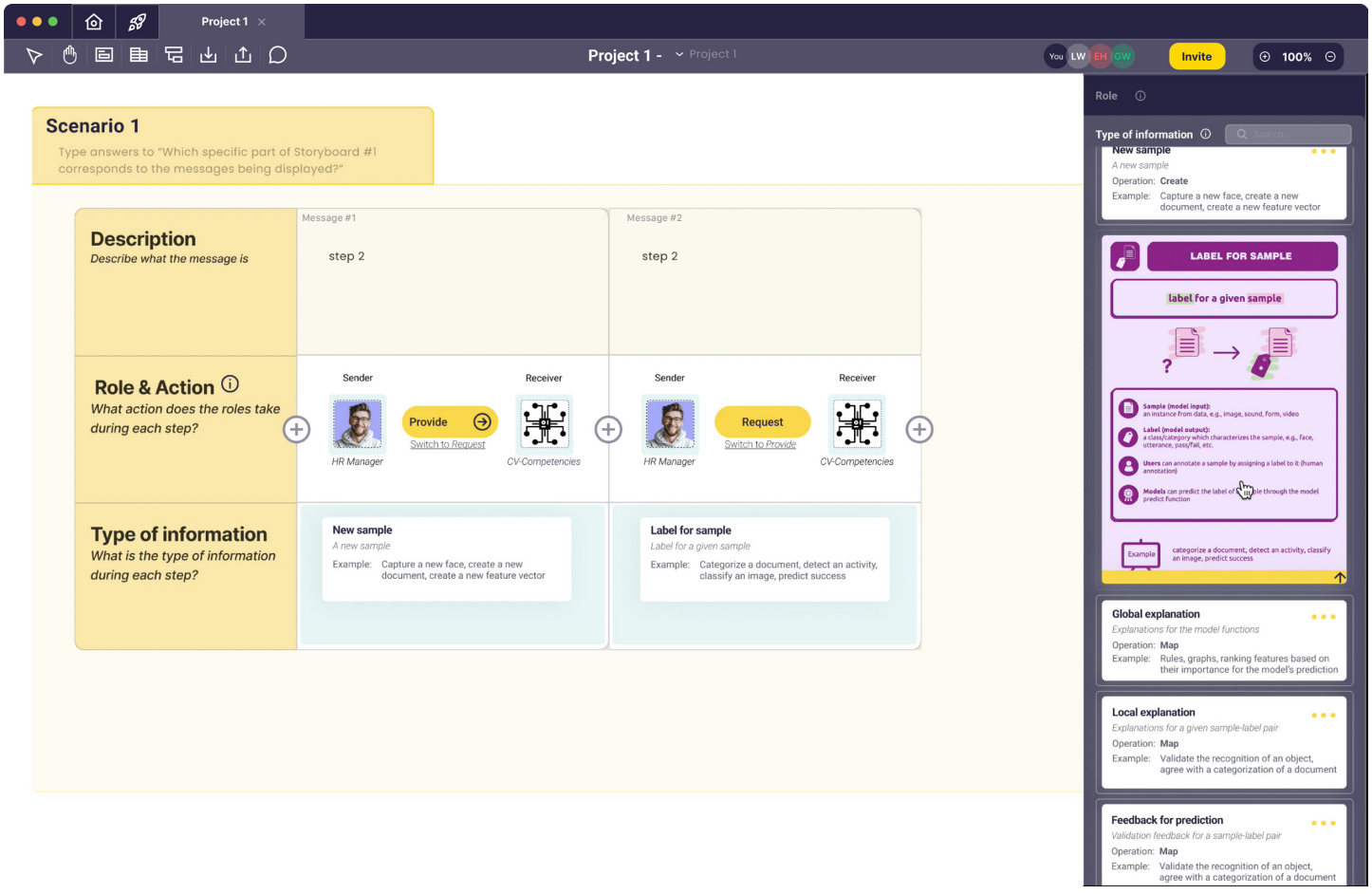


Figure 33.2. Details about Type of information on the right function part.

5.2.5 Interaction version

The Interaction interface can be accessed when a designer wants to connect non–adjacent information, create a flow of information outside of a linear structure, or view the actual cases corresponding to each piece of information. This is shown in *Figure 34.1*.

As you can see in *Figure 34.1*, there is a new Interaction Row at the bottom. This row includes Instance level, Data level, UI level, etc. The main purpose of this row is to visually present the figurative content of each message, i.e., the part that is relevant to the design case the designer is working on.

On the right, a workspace can be seen (as is shown in *Figure 34.1*). This will present the Interaction–related content of each Message, including the Type of Information, the connection routes that exist, the visualized content in the Interaction Row (named Data flow here), and so on.

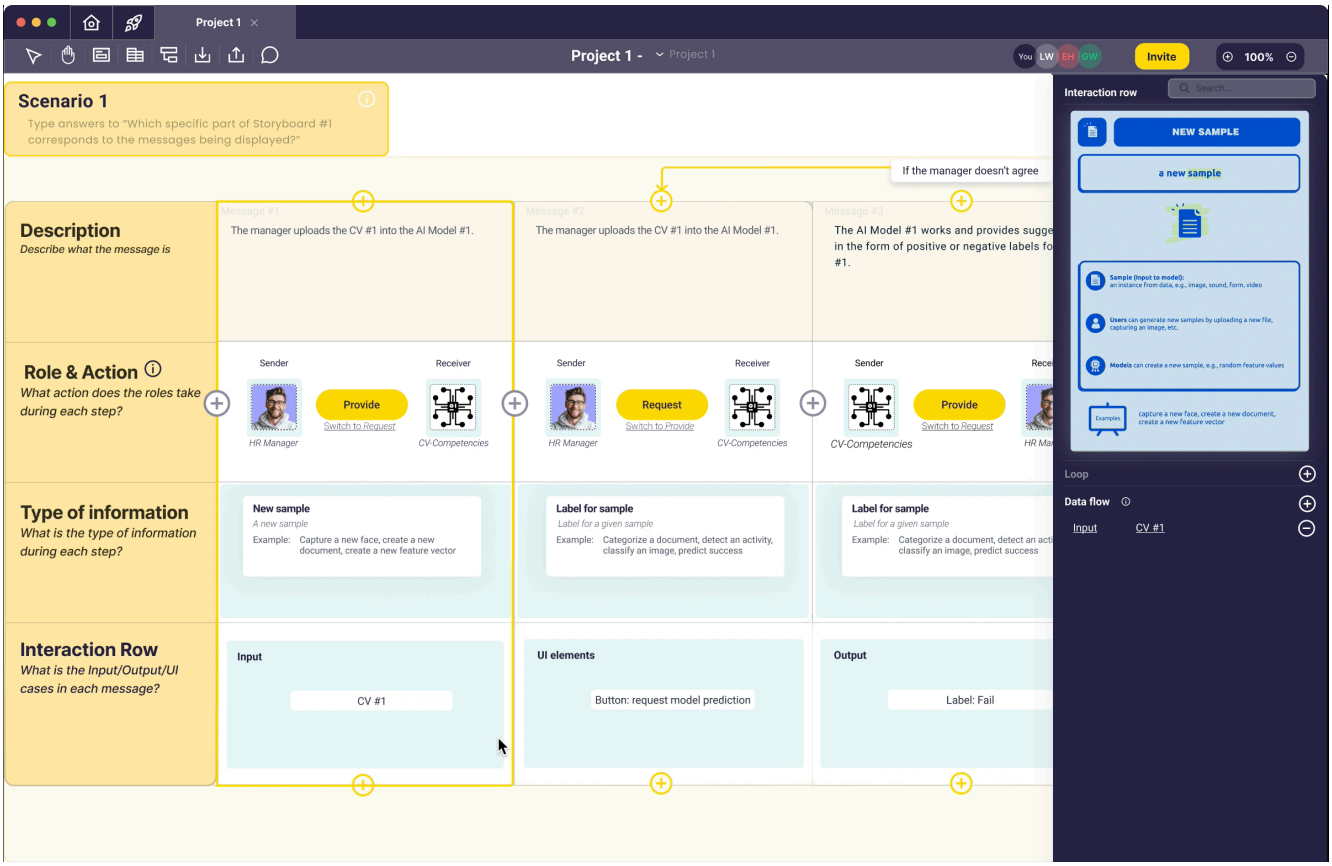


Figure 34.1. Example of Interaction version.

At the same time, two yellow "Add" buttons will be added to each message. This adds a non–linear message flow structure. As the connection between Message #2 & Mesage #4 in *Figure 34.2*. This means that if designers want to have interaction or data exchange between two non–neighboring messages, they can click on the yellow "Add" buttons to structure the direction of the message flow and the interaction.

It is important to note that the format of the Data flow can be limited or customized by the designer, with the main choice being related to the Type of information in the message. For example, the New sample in *Figure 34.3* gives two choices of Input/Output which are presented by the interactive element Call–out. *Figure 34.4* shows more choices for selection.

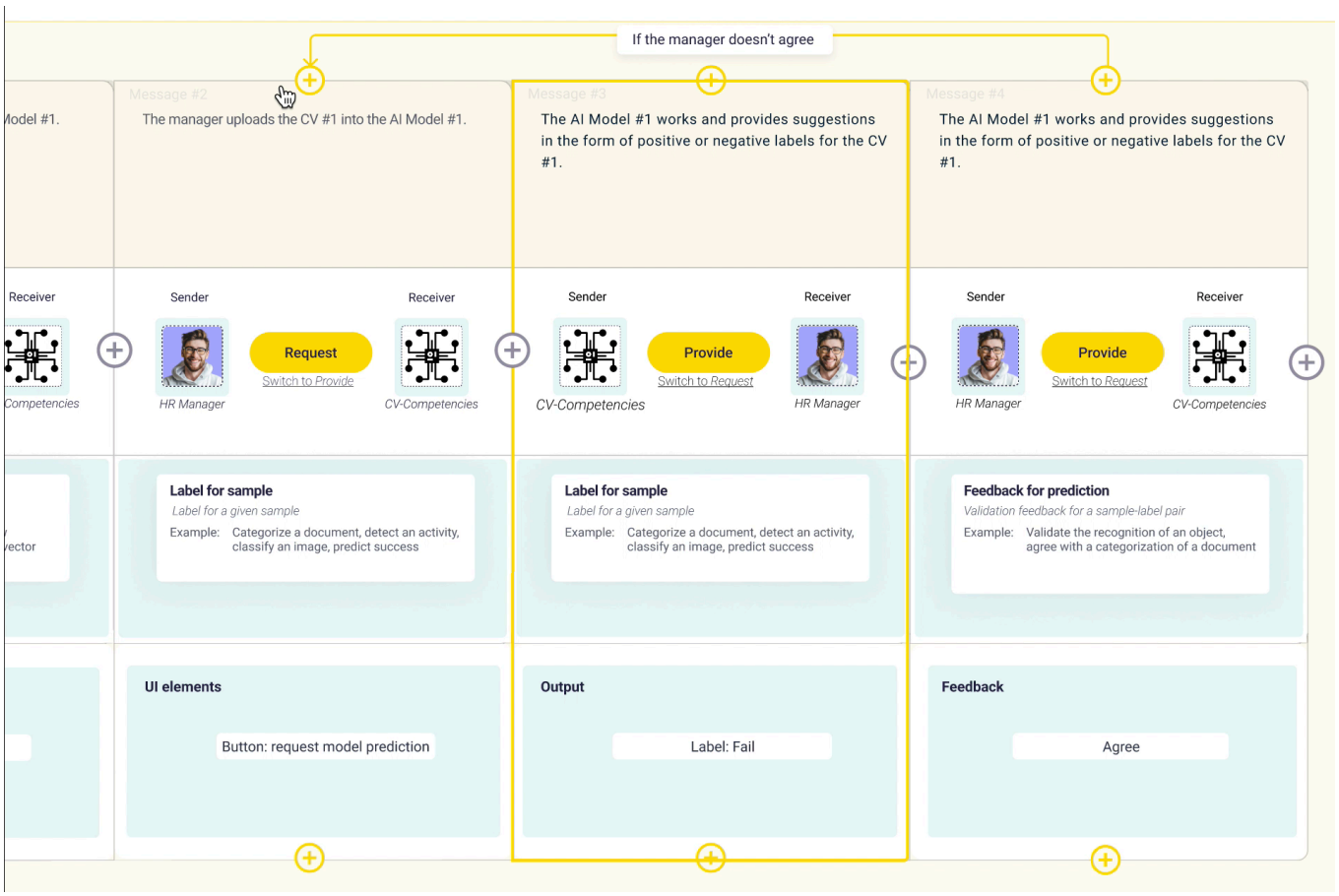


Figure 34.2. Connect messages for nonlinear structures.

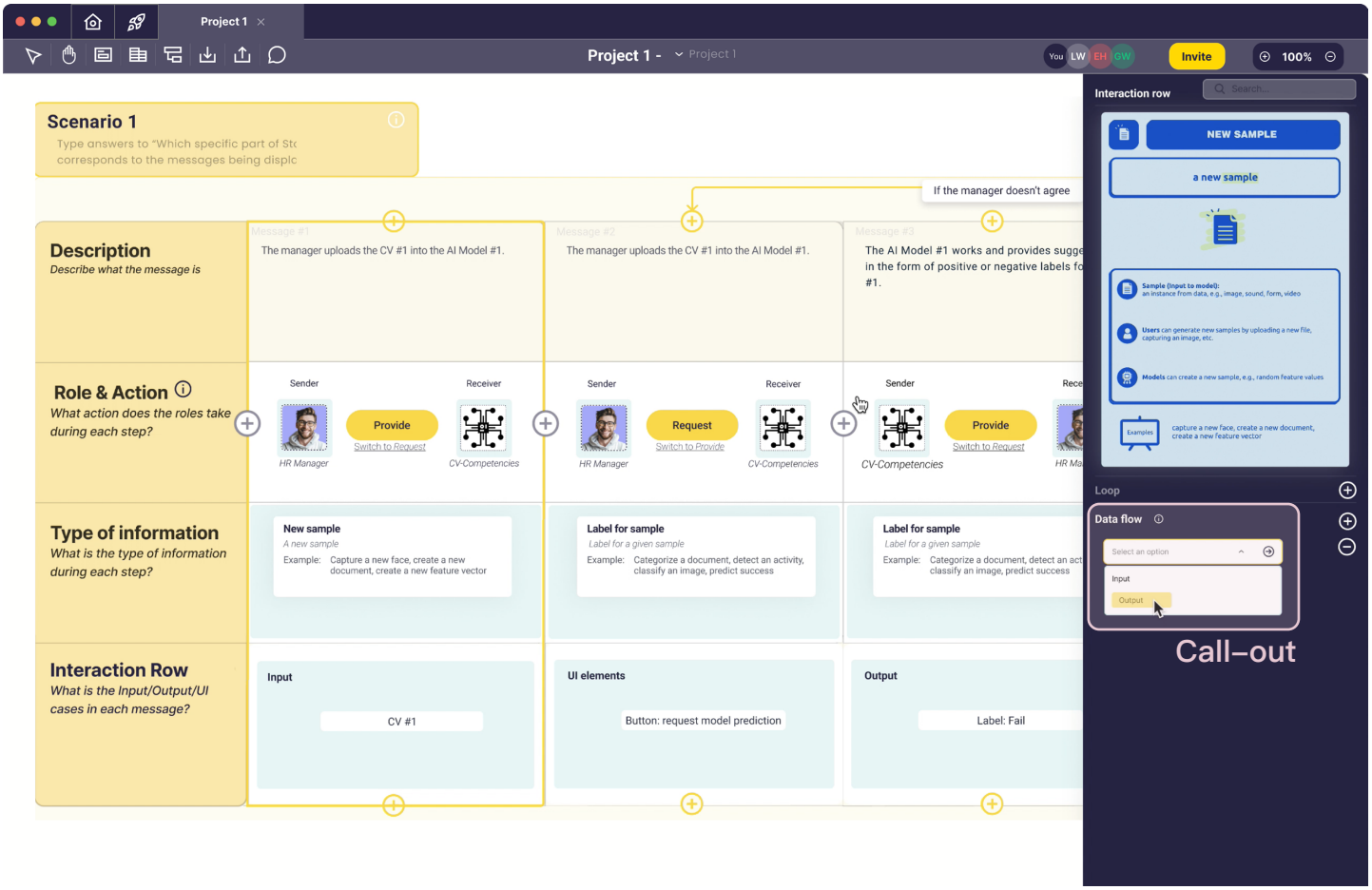


Figure 34.3. Example of choices for Data flow.

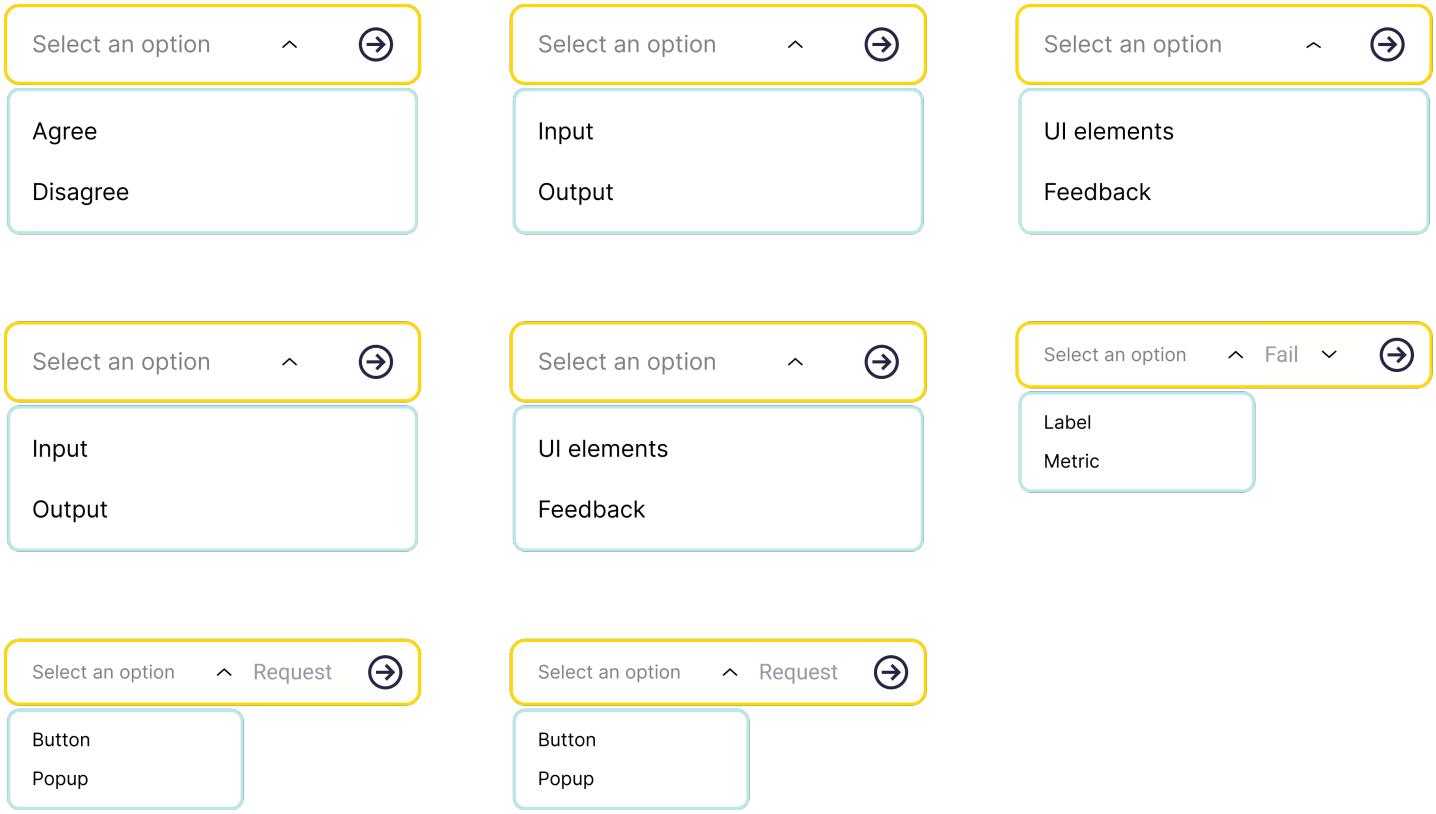


Figure 34.4. Choices for Data flow.

5.3 Test goals

After determining and modeling the primary functionality and interaction styles of the interface, the next step involves conducting functional testing of the design outputs. This choice is driven by three key reasons rather than opting for usability testing at this time.

Firstly, these features are structured and designed to align with the design goals outlined in Chapter Four, specifically targeting objectives like "**is easy to learn how to use the tool**" and "**Help designers learn the basic concept of 'Communicative acts' of Human–AI Interaction involved in this project**." Consequently, a test is necessary to evaluate whether the design outputs align with these expectations.

Secondly, although segments of the final design concept underwent limited testing with various inputs during the design phase, the design product, including interactions, has not been thoroughly tested. This raises uncertainty about whether designers understand the concepts and components through the interaction with the prototype.

Finally, despite having a interface based on html/css/javascript from the design process, it cannot support the test. The interactive interface built in Figma represents only a partial aspect of the complete interface, rendering it unsuitable for usability testing.

For more details about the test plan and the use case, see Appendix H.

5.3.1 Testable targets

Based on the previous section, it can be seen that there are two important goals of this test:

- 1). to assess whether the design outputs fulfill the mentioned design objectives: "is easy to learn how to use the tool" and "Help designers learn the basic concept of 'Communicative acts' of Human–AI Interaction involved in this project."
- 2). to assess whether designers understand the underlying concepts and components through the interaction with the prototype.

It is important to gain insights into these two test goals by how the testable objectives are set. Based on the functionality of the clickable prototype and the correlation analysis with the basic concepts and other elements involved in the project, the testable objectives for this test are as follows:

- – Participants recognize how to create, view and/or delete a new message/ a new role(human/model) independently.
- – Participants are able to create, show and/or delete the Input/Output/UI elements of each message.
- – Participants identify how Human–AI Interaction is described through messages, model–based information, and communicative acts.
- – Participants are able to find functions they need from the interface comfortably.

For the details of tasks set up in tests with testable targets, refer to Appendix H.

5.3.2 Data gathering method

The test was used in three test methods. Firstly, through interviews designed to get participants' responses to specific questions aligning with testable objectives. Secondly, by encouraging participants to "think aloud,"[45] the researcher noted key observations and subsequently addressed them in a follow–up interview. Lastly, the researcher directly observes the participant's behavior, documenting essential notes, and posing questions during subsequent follow–up interactions.

The "think–aloud"[45] method in user testing involves participants verbalizing their thoughts and feelings as they interact with a product or system. Participants express their reactions, opinions, and decision–making processes out loud in real–time, providing researchers with insights into their cognitive processes. This technique helps uncover user expectations, frustrations, and areas for improvement in the user interface or overall user experience. By listening to users' spoken thoughts, researchers gain a better understanding of how users perceive and navigate the system, which can inform design decisions and optimizations.

Here is an overview of the methods used in the test. See *Figure 35*.

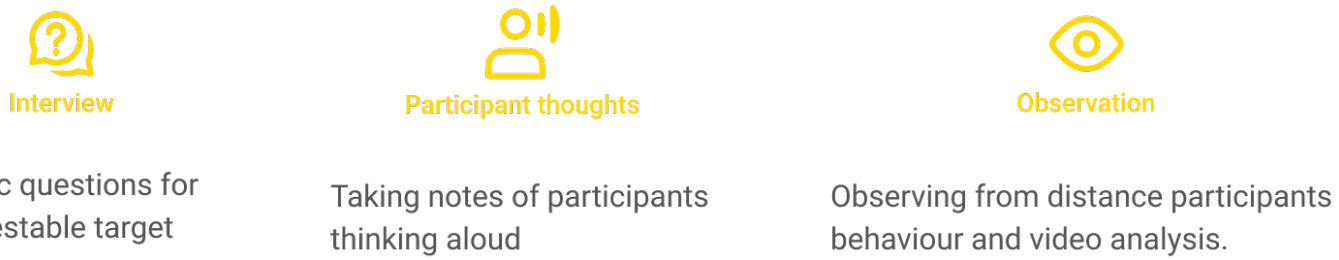


Figure 35. Test methods.

5.4 Test setup

This test was not only a functional test of the created interface, but also a test of the process, tasks, and reading materials for the final test. The final design outputs and final test content will be refined based on the results of this testing.

5.4.1 Preparation

To accomplish these goals, the test consisted of 7 steps (*Figure 36.1*) and 3 activity sections (*Figure 36.2*).

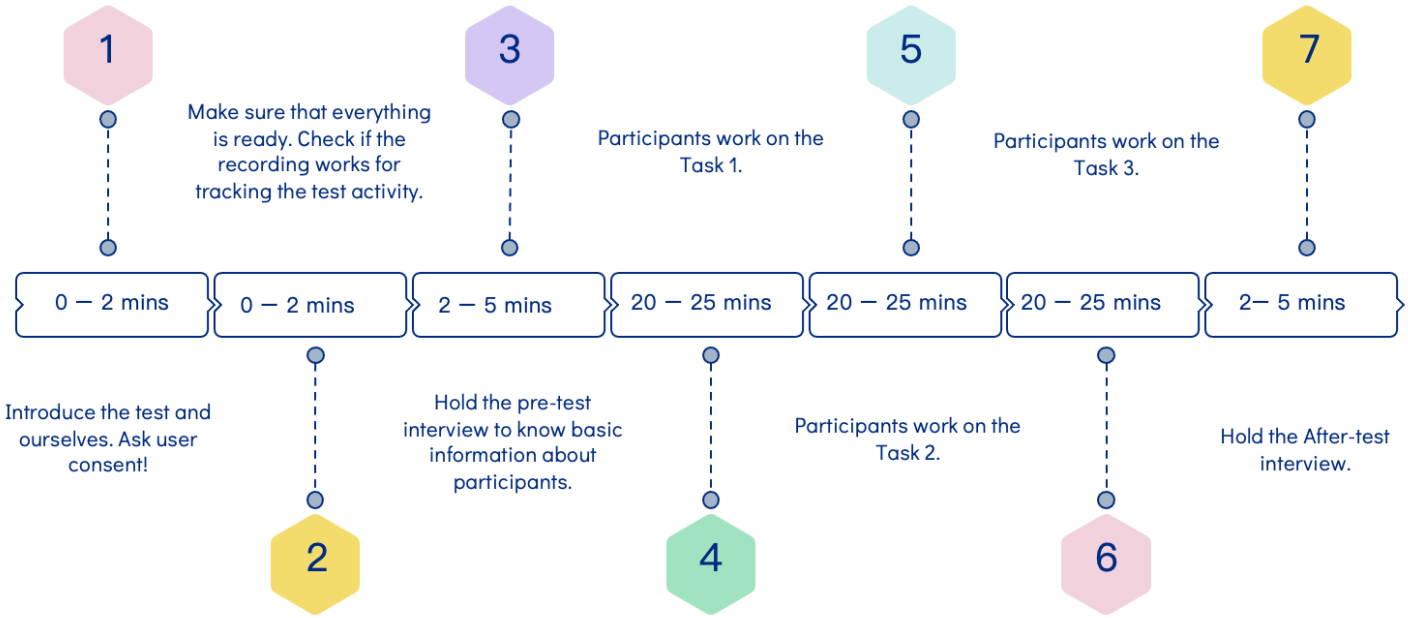


Figure 36.1. Overview of 7 steps.

The following is a detailed description of the goals of the 3 tasks, their content, and how they contributed to testable targets. The following is a detailed description of the purpose and content of these 3 tasks and how they contribute to the achievement of the testable objectives. It is important to note that this test is also the pilot test for the final test, so the overall task flow also needs to be tested and analyzed.

For more details about activities and materials, see Appendix H.

Test targets:

- – Participants recognize how to create, view and/or delete a new message/ a new role(human/model) independently.
- – Participants are able to create, show and/or delete the Input/Output/UI elements of each message.
- – Participants identify how Human–AI Interaction is described through messages, model–based information, and communicative acts.
- – Participants are able to find functions they need from the interface comfortably.
- – Assess whether the 3 tasks help find answers to the above questions, the test material easy to understand, and the digital prototype works.

Task 1

Goal: Assess if participants could understand concepts about communicative acts in the digital prototype.
Content: Questions here guided participants to explore by clicking the digital prototype, mapping the single Message to concepts in Communicative acts.
Contribution for testable targets:

Task 2

Goal: Assess if participants could understand concepts about communicative acts in the digital prototype.
Content: Following questions in the task, participants got educated on how Interaction row and communicative acts of the Human–AI Interactions worked in the digital prototype. They should map the use case with the Message sequence presented.
Contribution for testable targets:

Task 3

Goal: Assess if participants could understand concepts about communicative acts in the digital prototype. Assess the workflow and user experience of creating HAI based on the current digital prototype.
Content: Following the task description, Participants created, filled in and viewed a new message based on the use case. By doing that, they could experience the whole workflow of the digital prototyping based on knowledge got from the first 2 activities.
Contribution for testable targets:

The first task was to support participants' in-depth understanding of underlying concepts, including "Message," "Communicative acts," and "Human-AI Interaction." This was achieved by encouraging participants to explore the digital prototype using familiar design tools such as storyboards and personas. Through this exploration, participants gained valuable insights into these key concepts and effectively understood the content shown in the digital prototypes. The task questions provided prompts that guided participants in systematically exploring the interface. For more details about storyboards and personas, see Appendix H.

Before starting on this task, participants were exposed to the abstract concepts in the project. However, the examples embedded in the digital prototype, tailored to specific use cases, offered a more tangible representation of these abstract concepts. The participants' responses to the task questions served as crucial indicators, enabling an assessment of their recognization of the underlying concepts and their evaluative perspectives on the prototype.

The second task was enhancing participants' understanding of the complex project concepts. It is strategically divided the task into two activities to handle the complexity effectively, providing a well-balanced and supportive method to support participants in their understanding.

While the first activity centered on concepts related to "Message," the second activity showed the complexities of "Message" sequences, extending its scope to covering the content and interactions across the design considerations, including the "Message" level. Thoughtfully embedded guiding questions in the task, with supplementary materials such as storyboards, were designed to facilitate participants' exploration of the interface, thereby minimizing the overall learning workload. This intentional division and supportive approach aimed to maximize assistance for participants as they explored the complexity of the project concepts.

The final activity was strategically designed to assess participants' understanding of the project's concepts and evaluate the logicality of the digital prototype's functionality. Participants were tasked with creating new Message aligned with the given context, drawing upon their understanding of the basicl concepts and the Interactive row embedded in the digital model.

Functioning as a synthesis of the Task 1–2, this activity covered all six design considerations and their associated interactable features. Its successful completion was based on participants' well–understood of the overall concept and interactions established in the preceding tasks. To enhance the participants' experience and reduce the task's complexity, a new storyboard (*Figure 36.2*) was introduced. While maintaining consistency with the previous version, this updated storyboard included numbering in the lower right corner, aligning with the information in the digital prototype. This addition aimed to reduce participants' challenges in understadning the test's content and contextual elements. This task provided participants with a overal design experience, enabling them to bridge insights gained in the earlier tasks and offer valuable feedback during subsequent interviews.

Figure 36.2. The storyboard. Visualization elements [40]: Freepik.com

In testing, the prototype design used did not follow the ideal design above. This is because the ideal design has a large number of design elements that are not currently relevant to the creation of the HAI or the testing objectives, such as features like Add comments shown in *Figure 31.2*. Therefore, making the digital prototypes used for testing more intuitive and simple is necessary in order to reduce the cognitive load on participants during testing. Therefore, in the test, interactive content such as the toolbar, the top bar, collaborative boards, scale display, etc. were removed and only the core functionality was retained, as shown in *Figure 36.3*.

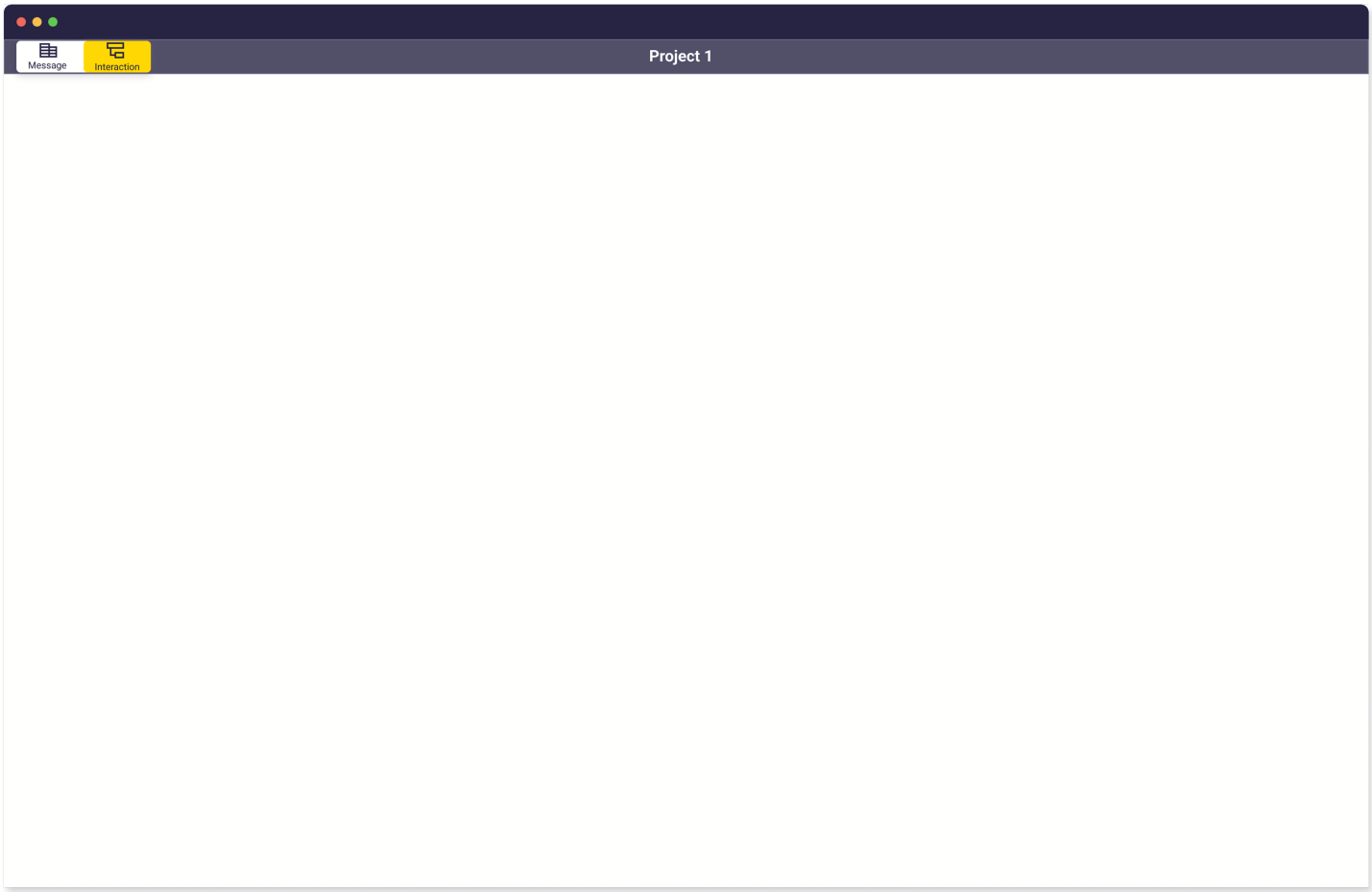


Figure 36.3. Blank interface.

There is one video visualizes part of the workflow of the model used in this test, showing the tasks that can be achieved with the prototype: 1) creating messages in the scenario; 2) filling in the blank messages; 3) creating touchpoints corresponding to the message content; 4) creating non–linear channels for the flow of information; and 5) creating the different personas.

The link: Video for Test 2

5.4.2 During the Test 2

A total of 2 participants were invited to take part in the Test 2 which also worked as the pilot test: they are both studying on Design for Interaction, from the TUD IDE Institute. The speech transcription files from the test and the content they created are shown in Appendix I.

Overall, participants were able to quickly understand and immerse themselves in the context of the task through posters, personas, and storyboards. The introduction of the project and digital prototyping background in the first stage also helped participants to understand the basic concepts involved in the project and try to apply them in the subsequent tasks.

The "think aloud" approach to exploration also maximized the researchers' ability to track participants' confusion or mental activity when using the digital prototypes, and provided an important opportunity to gain insights during the subsequent interviews.

5.4.3 Insights

The test serves two purpose: assessing the feasibility and identifying areas for improvement in the test plan while also collecting valuable design insights and paving the way for future development. Consequently, this part is structured to address insights related to the test plan's efficacy and those working for the design content, ensuring an overal examination of the pilot phase.

The analysis method used in the analysis process and brainstorming and summarizing the information from the transcribed text. The process is shown in Appendix I.

The first part is about some insights for the current test preparation.

1. The analysis of two pilot tests highlights the necessity for a more clear introduction.

After analyzing the outcomes of two pilot tests, it's evident that the current test plan lacks a sufficiently specific introduction to the prototype. This deficiency results in participants dedicating more time at the onset to understanding the workflow of the prototype.

The initial introduction about the digital prototype was modified to present its purpose in straightforward language, aiming to assist designers in building interactions for information exchange between humans and AI. However, experimental results indicate that this revised introduction still falls short of conveying the prototype's significance. Important questions, such as 'When do I use this tool?' and 'What results should I expect?' remain unanswered.

To address these problems, considering the varied roles of other tools in the design process is crucial. Therefore, enhancing participants' understanding involves describing, in the introduction, the primary design stage where the prototype is involved. This helped participants to develop mental expectations about the outcome of the prototype. See *Figure 37.1*.

"But I'm curious as to how the form of interaction I'm creating will be presented. Is there a button that when I click on it, it will run the overall message or will he only run specific messages? How will it render the interaction? " —Participant 1

"I almost forget the definitions although you introduced in the beginning." —Participant 2

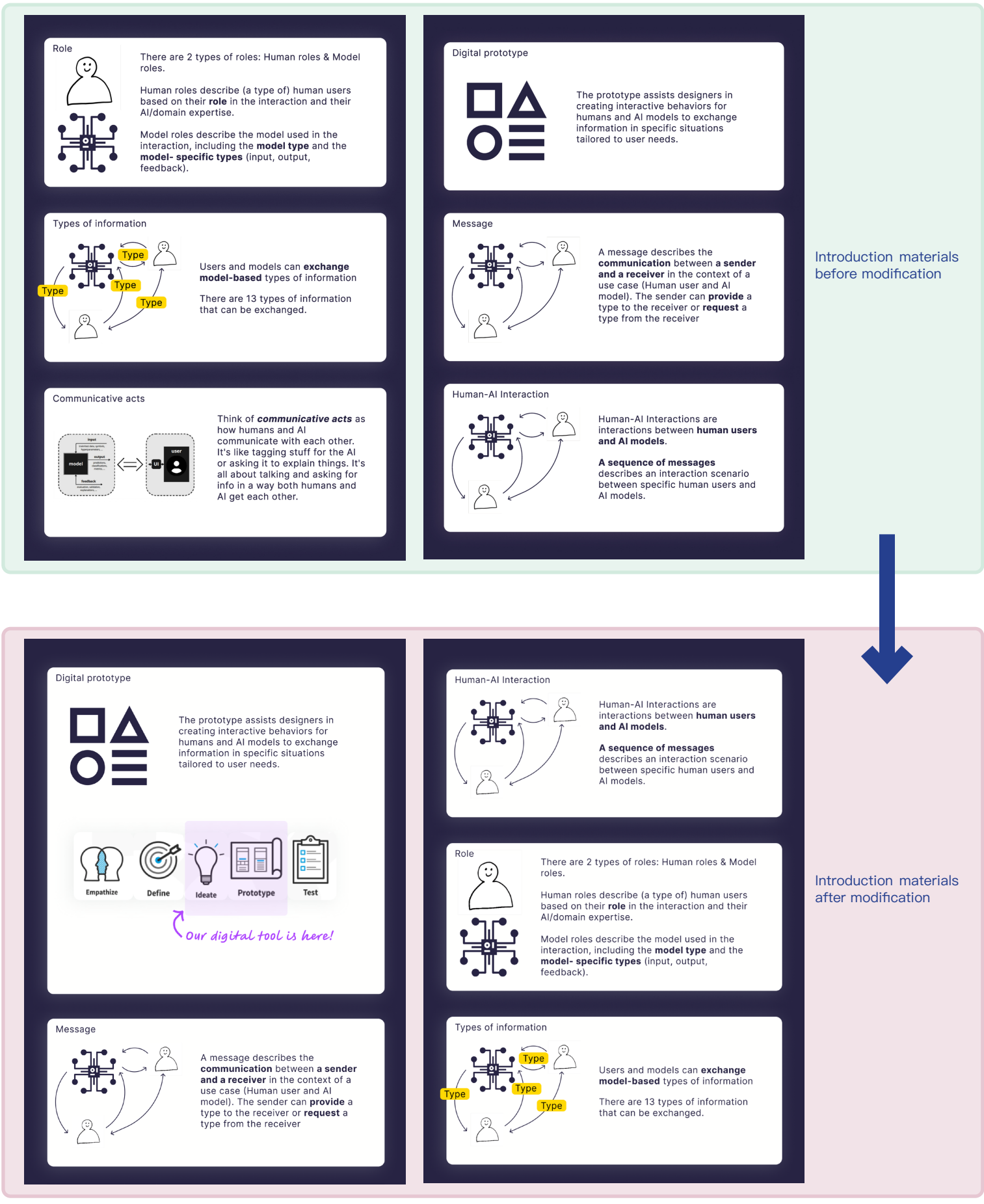


Figure 37.1. Introduction before modification & after modification.

2. Display all the parts that can be interacted with where they can be seen most easily, while ensuring that interactions do not interfere with task understanding.

In both tests, participants displayed high enthusiasm to create new blank messages; however, they tended to concentrate only on the front part of the message sequences, overlooking the interactive buttons at the back.

In task 1, it was observed that the functionality of the Provide/Request button could be interactive, but would change the meaning of Message. See A part in *Figure 37.2*. Considering the questions posed in Task 1 & 2, a potential improvement could be achieved by fixing this button, consequently eliminating the impact on the information presented in this particular case. See the A part in *Figure 37.3*. However, the button controlling the Provide/Request and being fixed in Task 1 & 2 could be clickable in Task 3

Meanwhile, participants in Task 2 could create a new Message by clicking the position of the button there, as shown in B part in *Figure 37.2*. Both participants were more preperred to click the first two buttons to see if they could add a message, failing to realize that the later buttons could be interacted with. Therefore, now the button to add a new blank Message is between Message #1 & #2. See the B part in *Figure 37.3*.

"Is there any button to add a new blank message? Ok. Maybe it would be easier to notice it if you put it in the first or second plus button." — Participants 2
"Is it (Type of informations) a button? For me, it is not like a button, because there is one icon nearby. And it's not in a location that's easy to notice. " —Participant 2

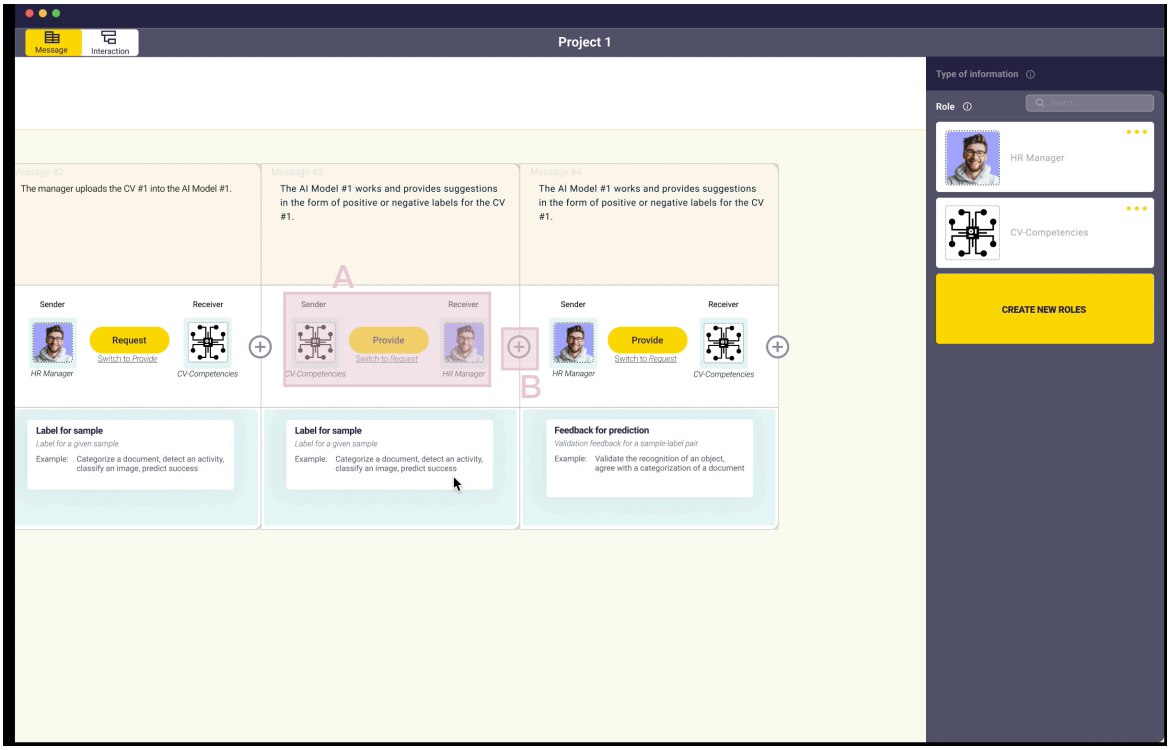


Figure 37.2. The interactive buttons in the prototype during the Test 2.

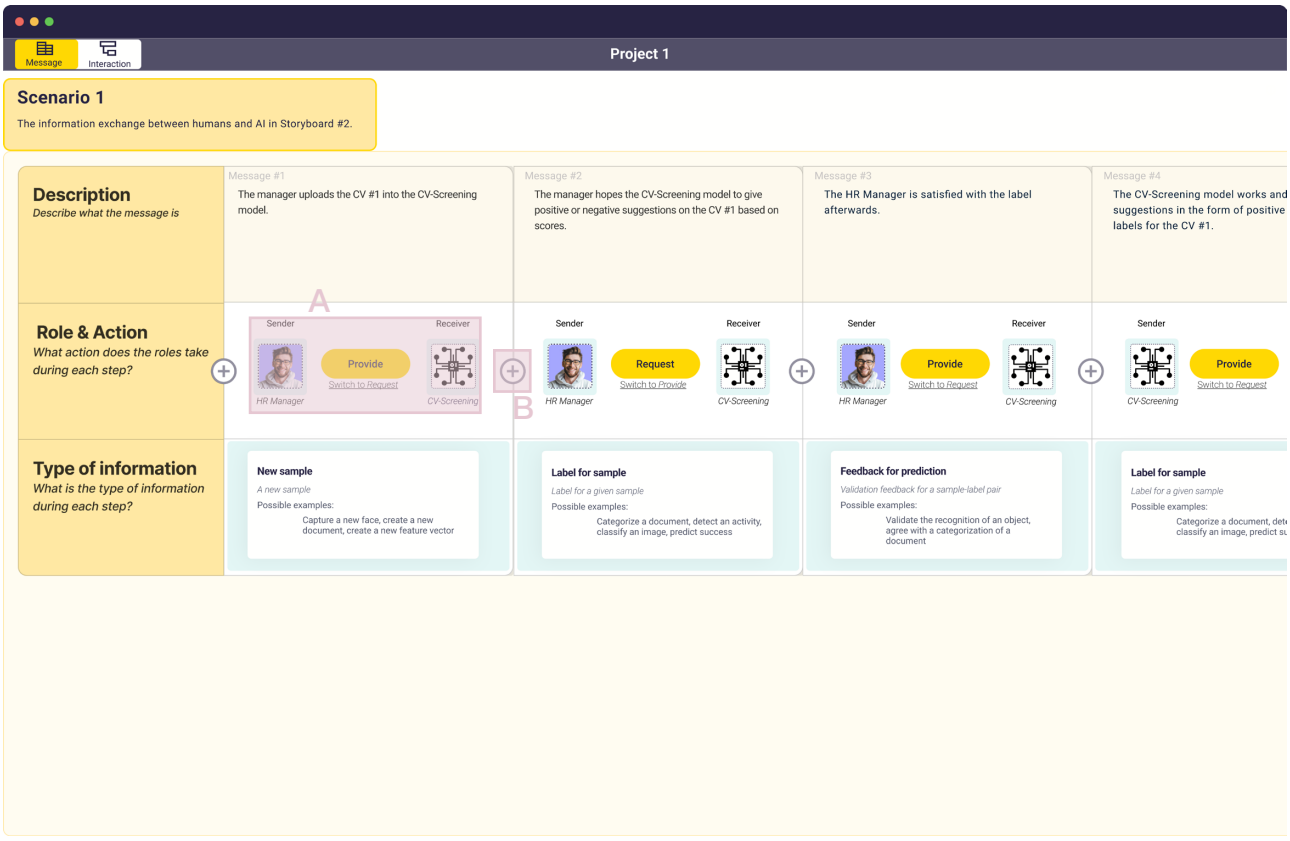


Figure 37.3. The interactive buttons in the prototype for the final test.

3. Reorganize the questions in each task so that they encourage participants to interact with the prototype.

As the second participant indicated, some of the overly detailed questions would cause her to spend too much time focusing on the details of the message at the expense of the totality. In the first two tasks the participants were mainly familiarized with the interaction with the prototype, so the design of the questions needed to be reconstructed to remove the overly detailed questions.

Also the current task description only required participants to understand what they were seeing and did not guide them to interact with the prototype. First, tasks have changed a lot. To address this, the new task descriptions now clearly guide participants to interact with the model. Instead of only "observing" the model, participants are encouraged to actively interact with it.

Here are some examples on new tasks. See *Figure 37.4*, one subtask was seeking details for a specific type of information hidden in another information layer. For *Figure 37.5*, one subtask was correcting content in Touchpoint. For *Figure 37.6*, the main Task 3 does not change a lot but the storyboard were changed slightly to provide participants with a more clear and easier tool to understand design task in the "CV–screening" use case. In detail, the storyboard have been reworked to focus more on the behavioral details of the content. At the same time, the task scenarios that needed to be added in Task 3 were added directly to the storyboard, making it easier for participants to understand.

For more details about the test materials for the final test, see them on Appendix J.

"Another reason why I wasn't sure about message before was the second question. The second question as I understand it is for me to decide by myself. I thought the task was subjective." — Participant 2

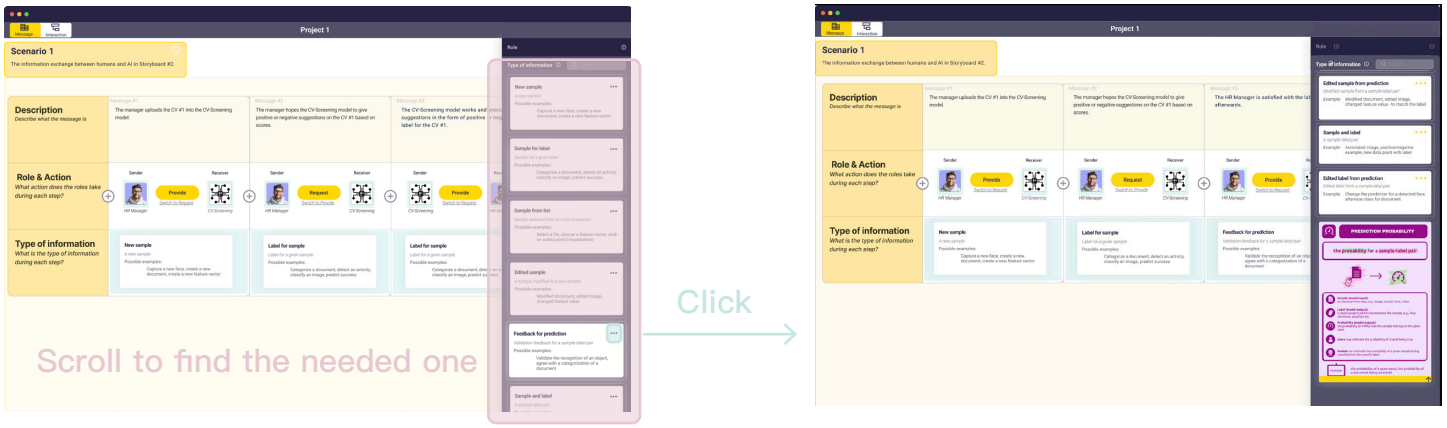


Figure 37.4. Example of the new Task 1.

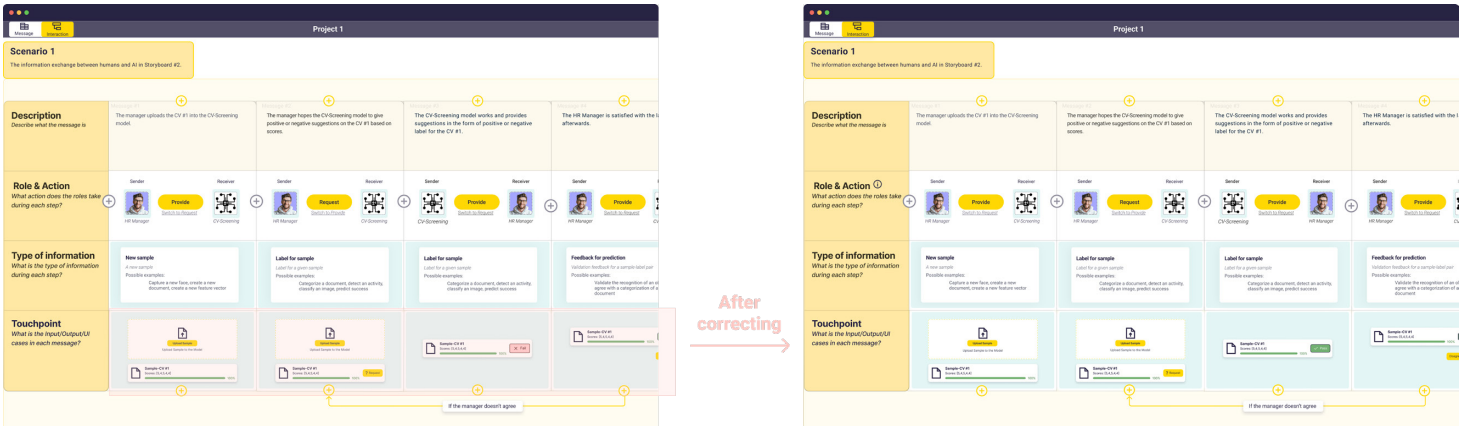


Figure 37.5. Example of the new Task 2.

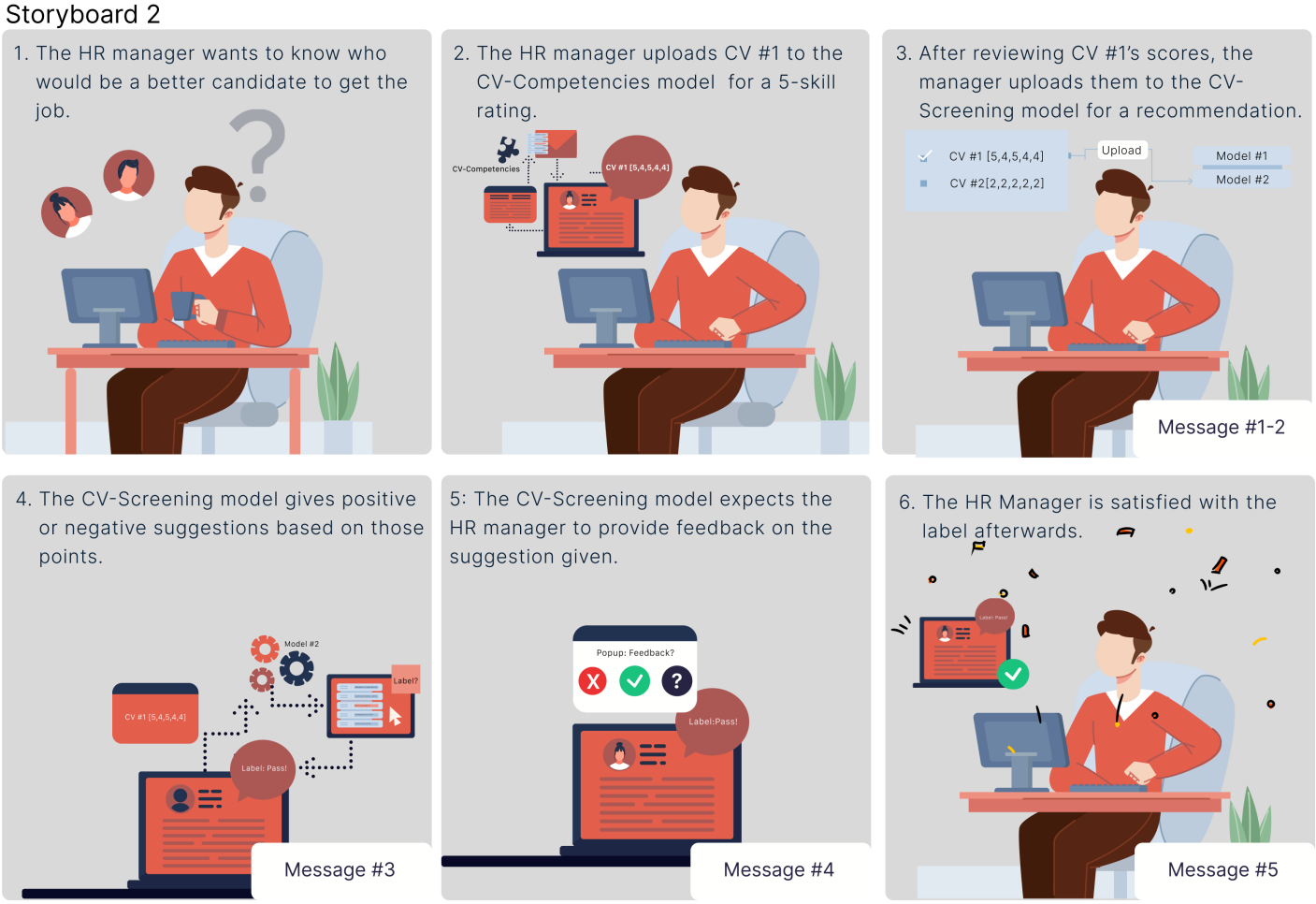


Figure 37.6. Add the responding task into the storyboard. Visualization elemetns [40]: Freepik.com

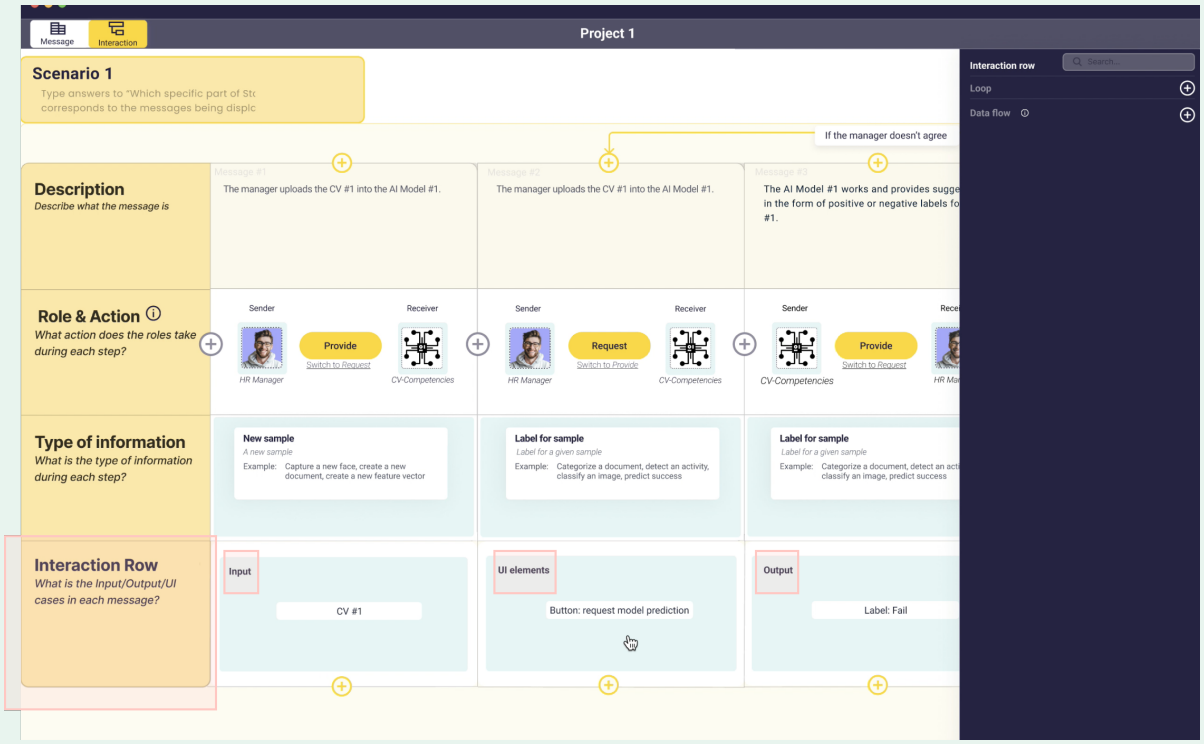
The second part is about some important insights for the design development. For more details about changes, see Chapter 6.

1. Change some descriptions in the prototype.

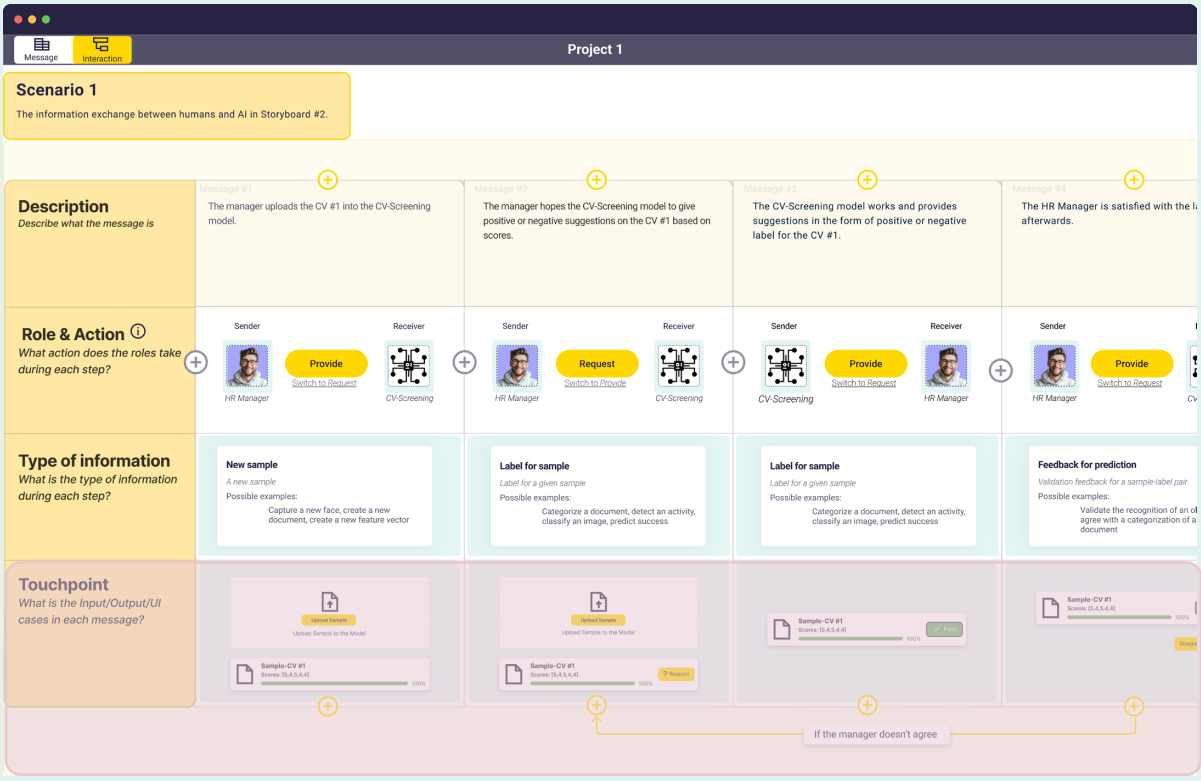
The prototype that was tested contained a number of terminology or interaction errors, and all interactions were used in the Pilot tests, documented found, and corrected.

Also changed "Interaction Row" to "Touchpoint". This better summarizes the relationship between Input/Output/UI elements and allows different levels of content to be presented at the same time. At the same time, Touchpoint is also a keyword used in user journey maps [47], which is easy for designers to understand. See *Figure 38*.

"Ok, I get it! It is something like the Touchpoint in the user journey map, right?" —Participant 2
"But if I were to do other use cases independently, I would be confused by the words Input/Output/UI elements. Because they give me the impression that they don't belong to the same category." — Participant 1



Interaction Row before modification



Touchpoint after modification

Figure 38. Example of words which should be changed.

2. Keep interactions for the same purpose the same and logical.

It is worth noting that there are inconsistencies in the interactions in the prototype. This means that in order to obtain the same purpose, participants may find 2 different interactions.

For example, see *Figure 39.1,* since the dark-colored function part was always fixed to the right and defaults to displaying Roles. As a result, when a participant clicked on a blank "Receiver" or "Sender", the content on the right didn't appear to change, whereas when they clicked on a blank "Type of information", the dark-colored function part on the right changed content. It not only defeated participants' expectation, but also led to a lack of uniformity and logic in the interaction.

"If I click the blank square of Sender or Provider, there is nothing changing on the right part. However, if I click the blank square of Type of information, the right function part would change. There is no uniformity in these two operations." —Participant 2

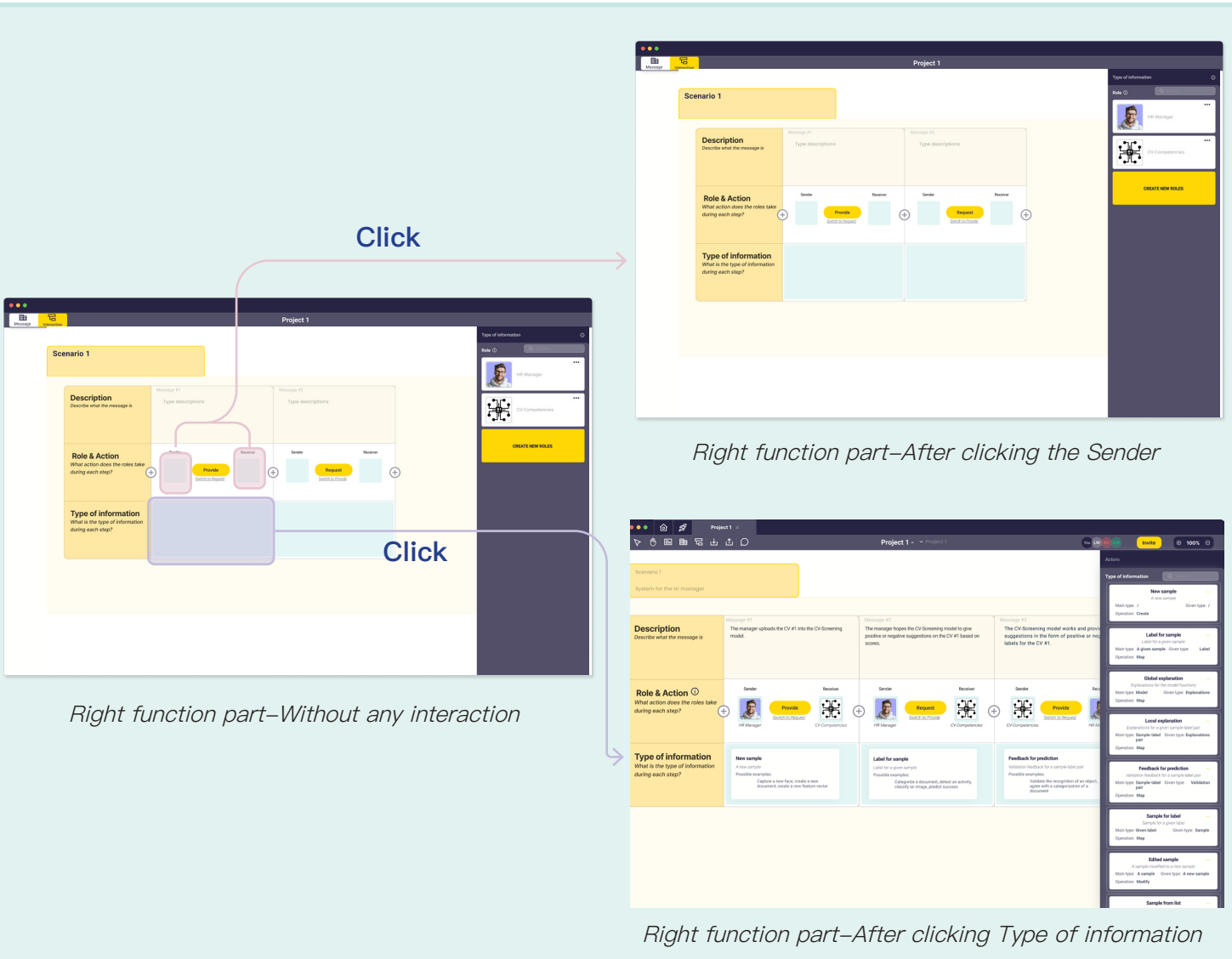
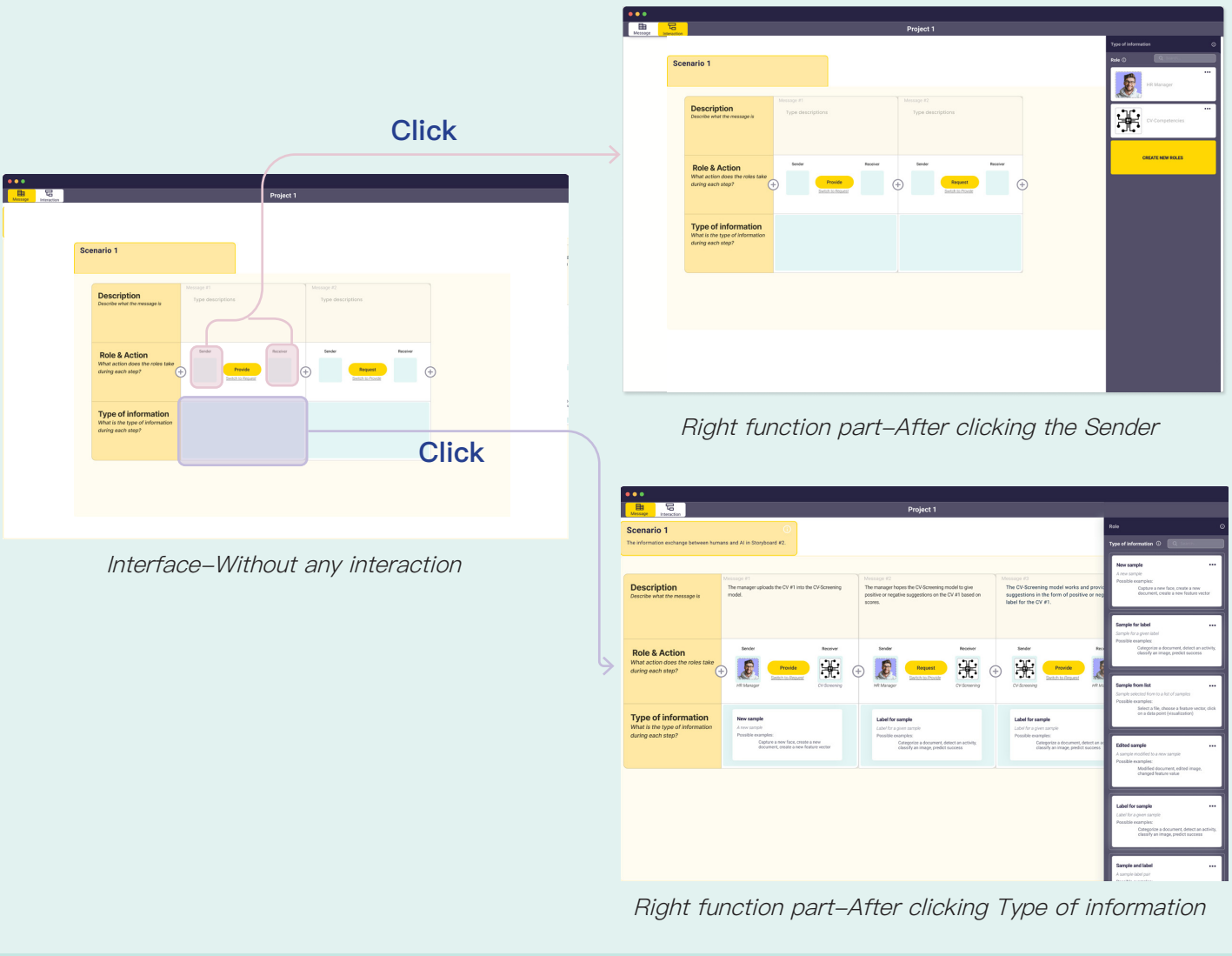


Figure 39.1 Example of the uniformity in the prototype for the Test 2.

Based on these insights, the final prototype developed. See *Figure 39.2.* It shows the final prototype hides the dark-colored function part if there is no interaction with the content in filling in the Receiver, Sender and Type of information. What's more, the new design solution brings more other benefits, such as providing a more immersive reading of Messages and Message sequences.

For more details about these changes, see Chapter 6.



Interface–Without any interaction

Right function part–After clicking the Sender

Right function part–After clicking Type of information

Figure 39.2 Example of the same interactions for the final prototypee.

3. Suggestions on UI components for the prototype.

Moving on to the UI problem, attention is drawn to the Type of Information details function part (the dark part fixed on the right). Some current buttons in this interface were found to be hard to capture participants' attention and triggering desired interactions. See *Figure 40.1*. Participants could click these yellow buttons to get details and explanations on types of information, however, it was hard to notice and click them in the current prototype. This issue may stop the seamless flow of user engagement and block the intended interactive experience. It has been developed in the final prototype by changing colors and icons. See *Figure 40.2*.

A second and equally critical UI problem centers around certain buttons that, according to users' feedback, fail to effectively convey information about their interactive potential. This ambiguity leaves participants uncertain about the actionable elements within the interface, diminishing the overall usability of the system. See *Figure 41.1*. Compared with “Interaction row” which cannot be clicked and “Data flow” which could be clicked, it was hard to distinguish. For one participant, with the word “Data flow” and the icon, it seemed to be unclicked. And it has been changed into something like in *Figure 41.2*.

In light of these UI challenges, prioritizing a overal review and potential redesign of button functionalities, visual cues, and overall interface clarity is needed. By refining these elements, we aim to optimize the user experience, fostering a more intuitive and efficient interaction with the interface. See more details about UI components change in Chapter 6.

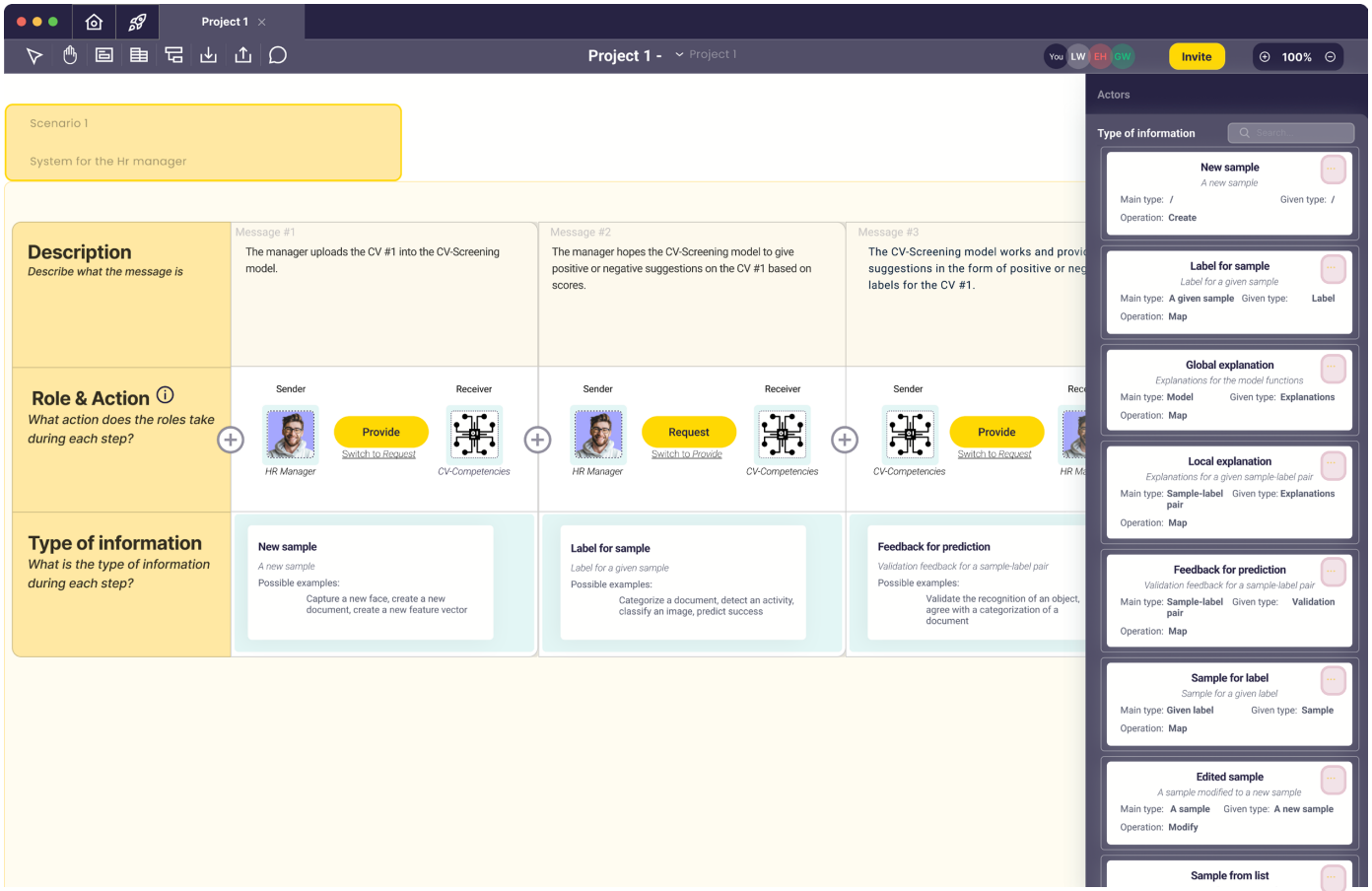


Figure 40.1. Example for buttons which were hard to be noticed.

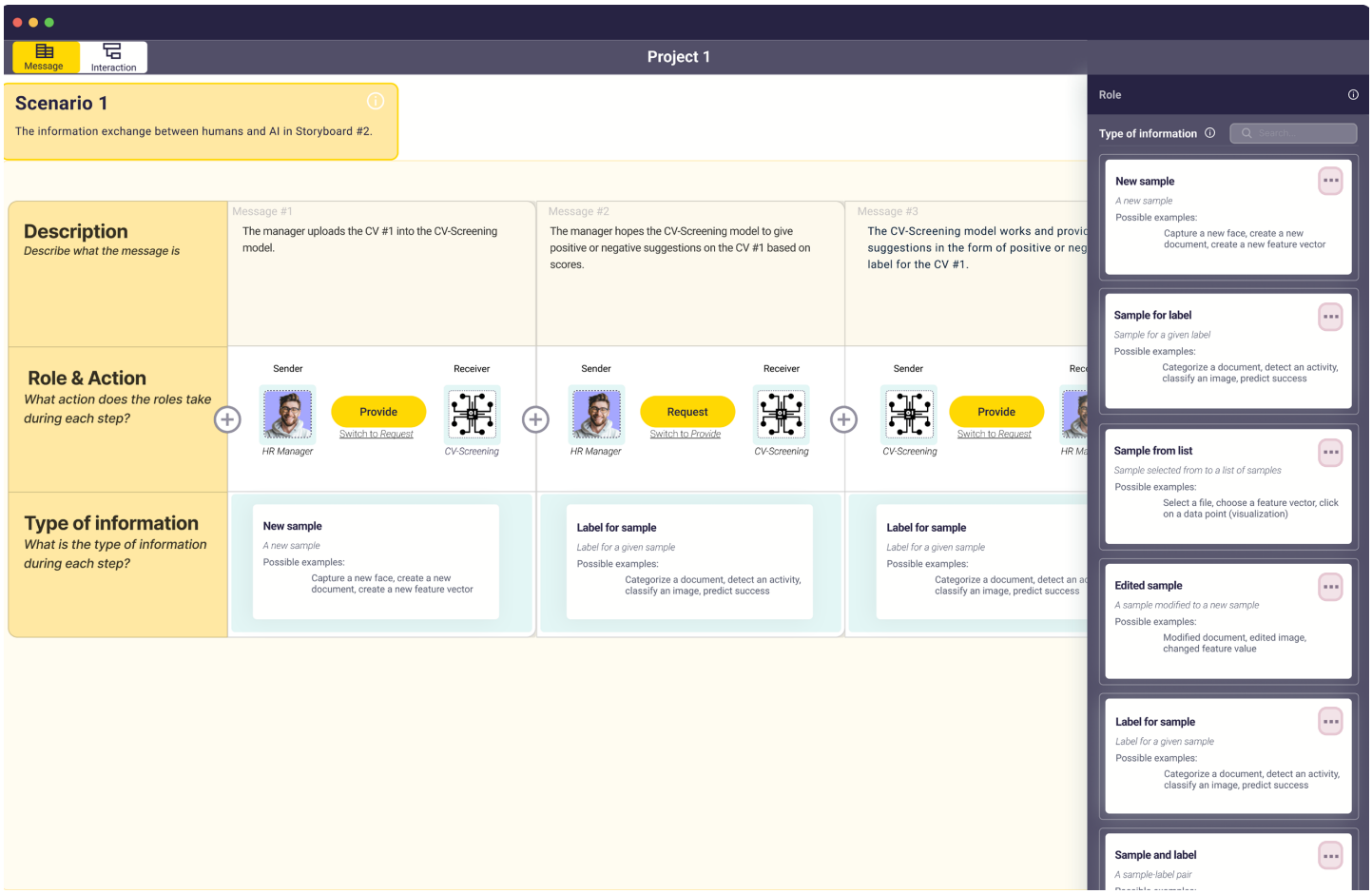


Figure 40.2. Example for buttons which are changed.

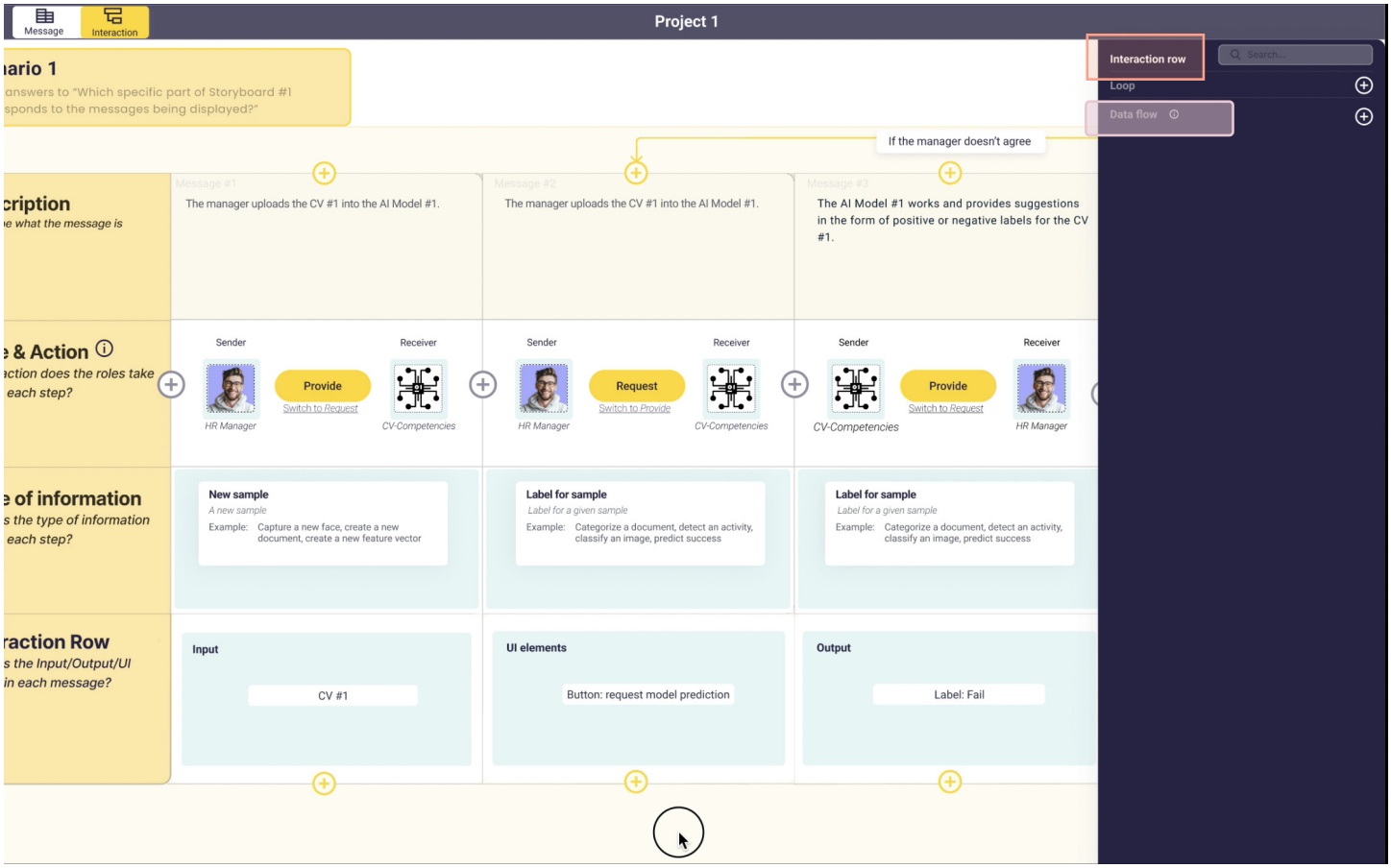


Figure 41.1. Example of distinguishing buttons which could be clicked and unclicked.

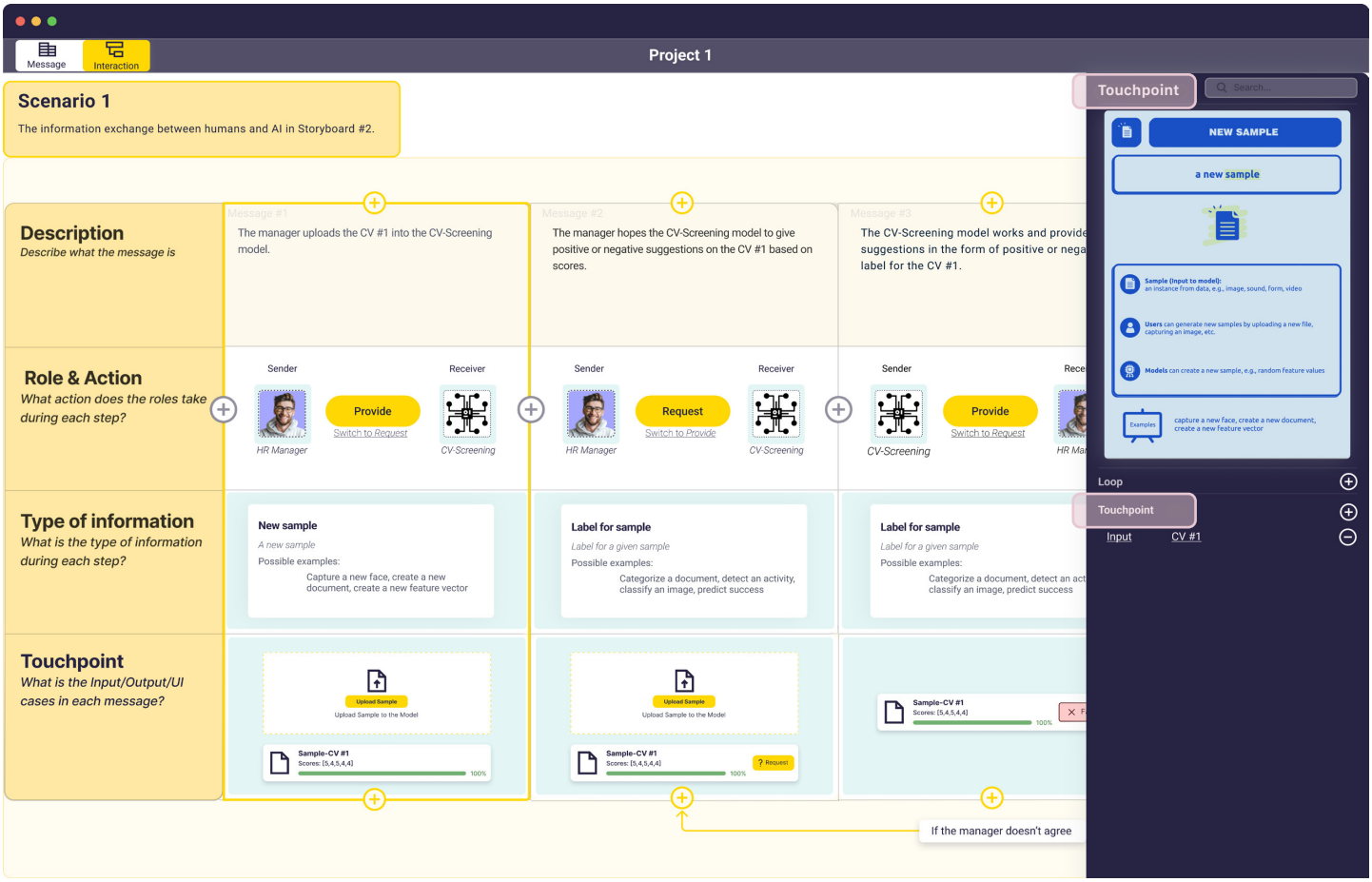
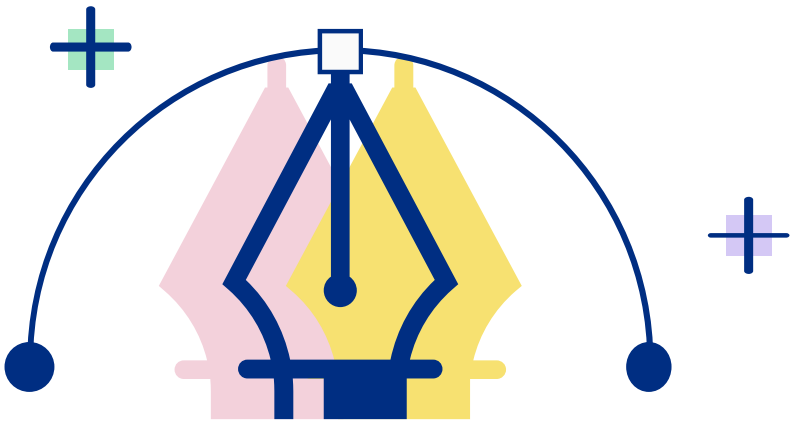


Figure 41.2. Example of distinguishing buttons which could be clicked and unclicked.



6. Final design & test

The sixth session will show the evaluative phase of the final design, including both the final design and the final test. This section also provides an in-depth look at the research activities carried out throughout the design-testing process and presents the conclusions drawn from the concluding analysis.

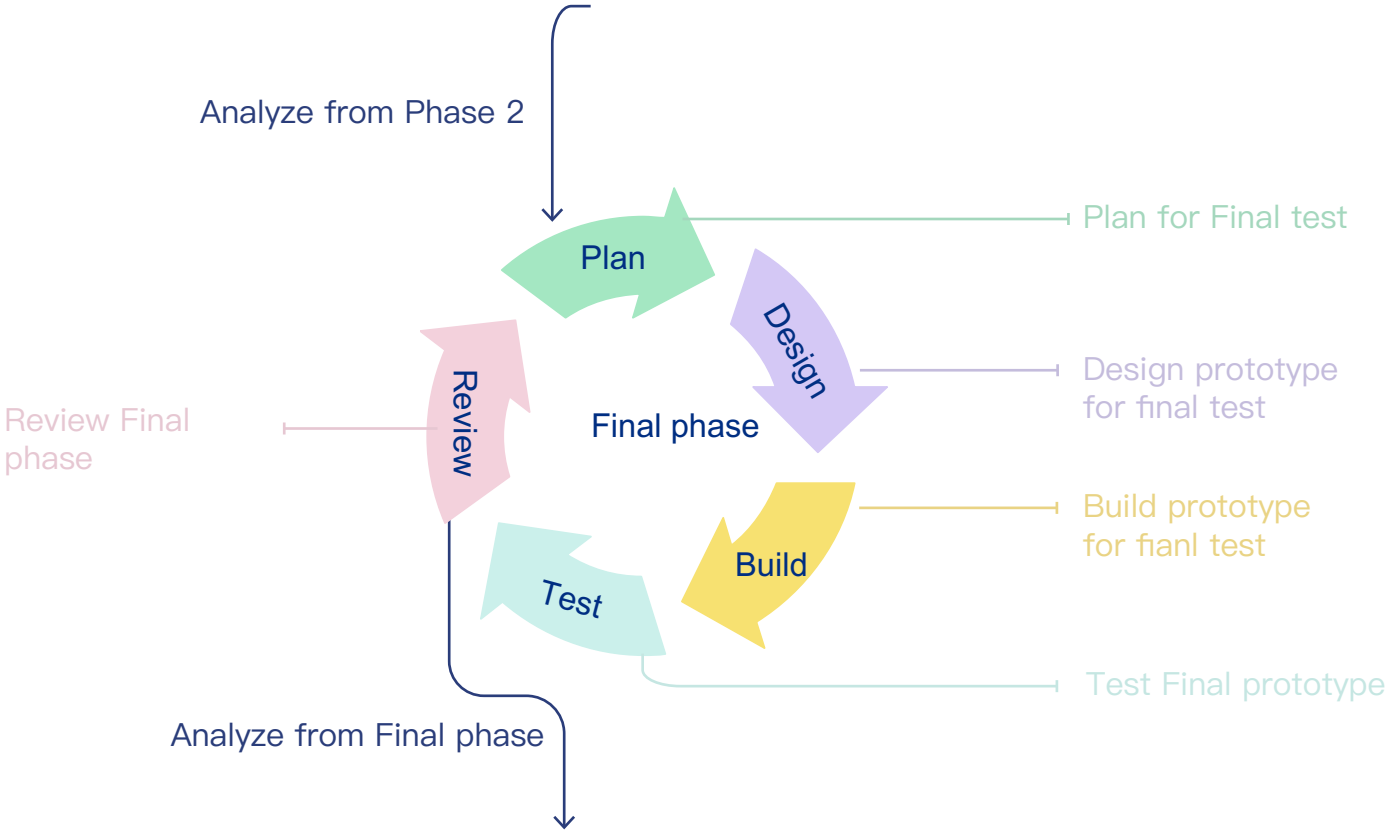


Figure 42.. Overview of Final phase.

6.1 Final design

The final output of the project is a digital tool designed to support the early stages of human–AI interaction design. Based on communicative acts and human–centered design, this tool assists designers during the Ideation stage of design thinking. It achieves this by visualizing the roles, data, and information involved in the process of information exchange during human–AI interactions. The aim is to enhance efficiency and ease in designing these interactions. This tool serves as a bridge, helping a seamless connection between design and AI exploration.

6.1.1 Overview

This design output was improved based on the design in Chapter 5 and the design insights gained from subsequent testing, with four main improvements:

1) simplification of the functionality and the information contained;
2) name and content of Touchpoint;
3) the reading experience of the Types of information details;
4) a more immersive reading Message (sequence) experience by hiding and interactively waking up the different functional areas on the right hand side.

Figures 42.1 & 42.2 represent the final design of the interface, which retains the two primary functions: the creation of the Message sequence and the interaction with the content in the Touchpoint section.

In the Insights–related content in Chapter 5, it has been discussed how to modify some of the test content, design concepts, and digital prototypes based on the insights obtained. More details about the design and test materials are provided next, while the unchanged tools and interactions remain the same as in Chapter 5.

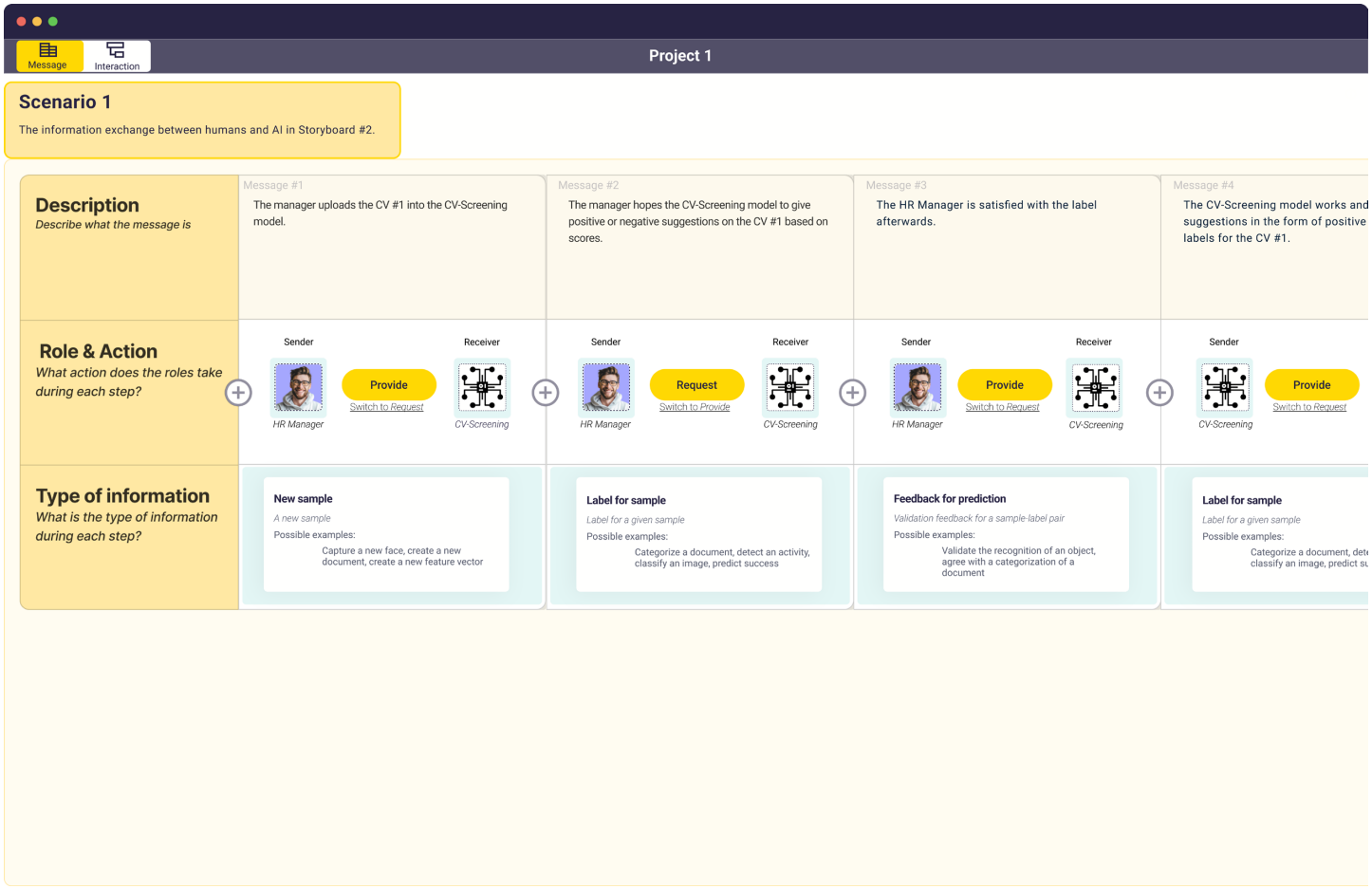


Figure 42.1. The interface for Message.

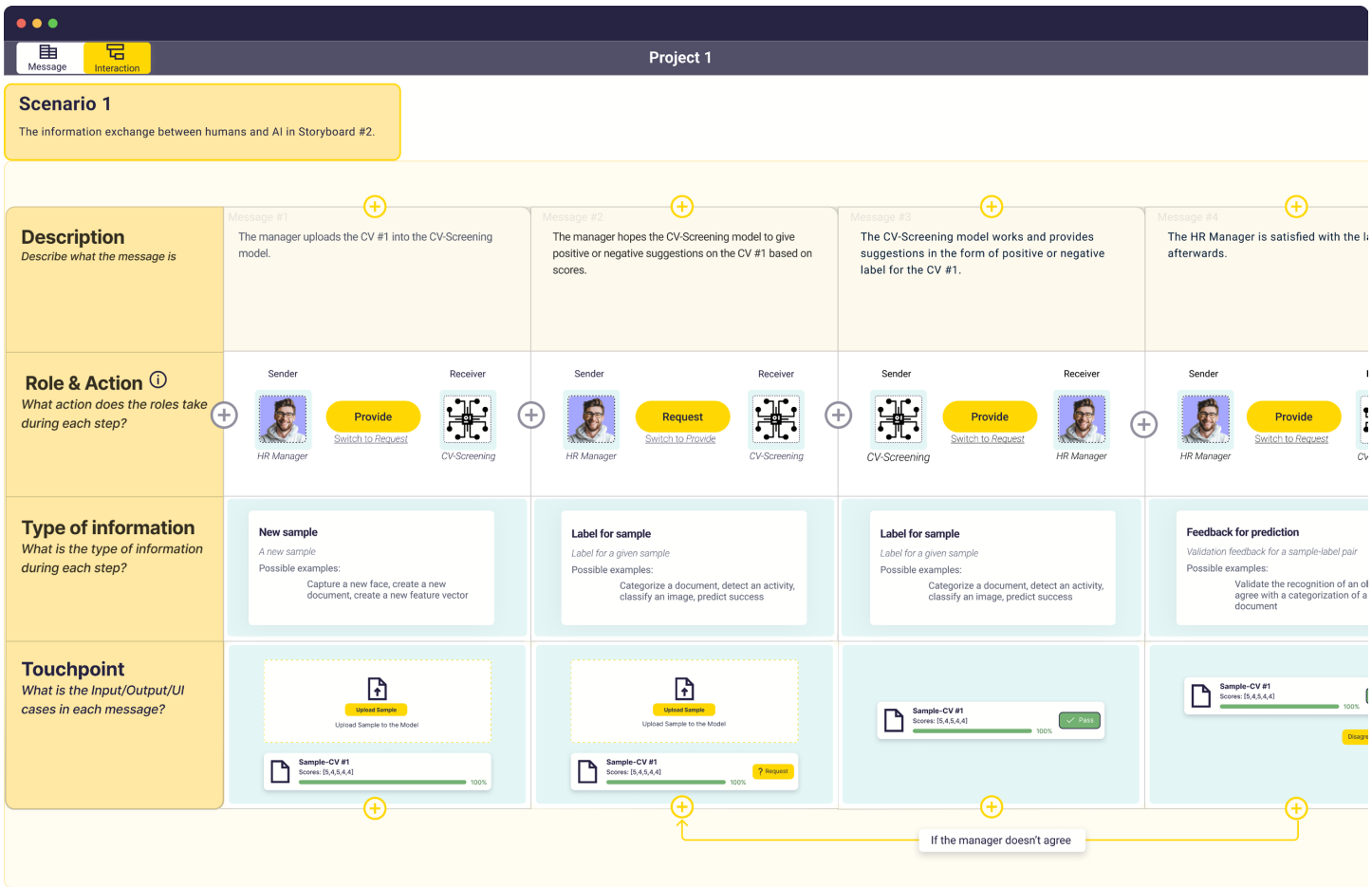


Figure 42.2. The interface with Touchpoint.

6.1.2 User interfaces

1. Top positioning status bar.

In the final design, the topmost function or display had been temporarily removed, which was the same as the prototype in the last test. This decision was made to prevent overwhelming participants with too many icons that don't serve any interactive purpose, potentially causing confusion. The elimination of non-interactive content aims to help participants focus on the main functionality of the test. See *Figure 43.1*.

Figure 43.1. Toolbar.

"Message" is the button to go to the interface for Message. It is shown in *Figure 42.1*. Designers could modify Messages in this interface. "Interaction" is the button to view and modify Touchpoint and connections of Messages. See *Figure 42.2*.

2. Project name.

On the middle of the second level, it is shown the name of the current project. But for the prototype, participant cannot type the name freely. See *Figure 43.2*.

Figure 43.2. Project name.

3. Collaborative boards & Scale display.

In the current prototype, here is no place for Collaborative boards & Scale display. The functions might be developed in the future.

4. Bench

This is the area that will be displayed after the Scenario and Message (sequence) are created. This is an infinite canvas and can be freely moved or zoomed in and out. See *Figure 43.3*.

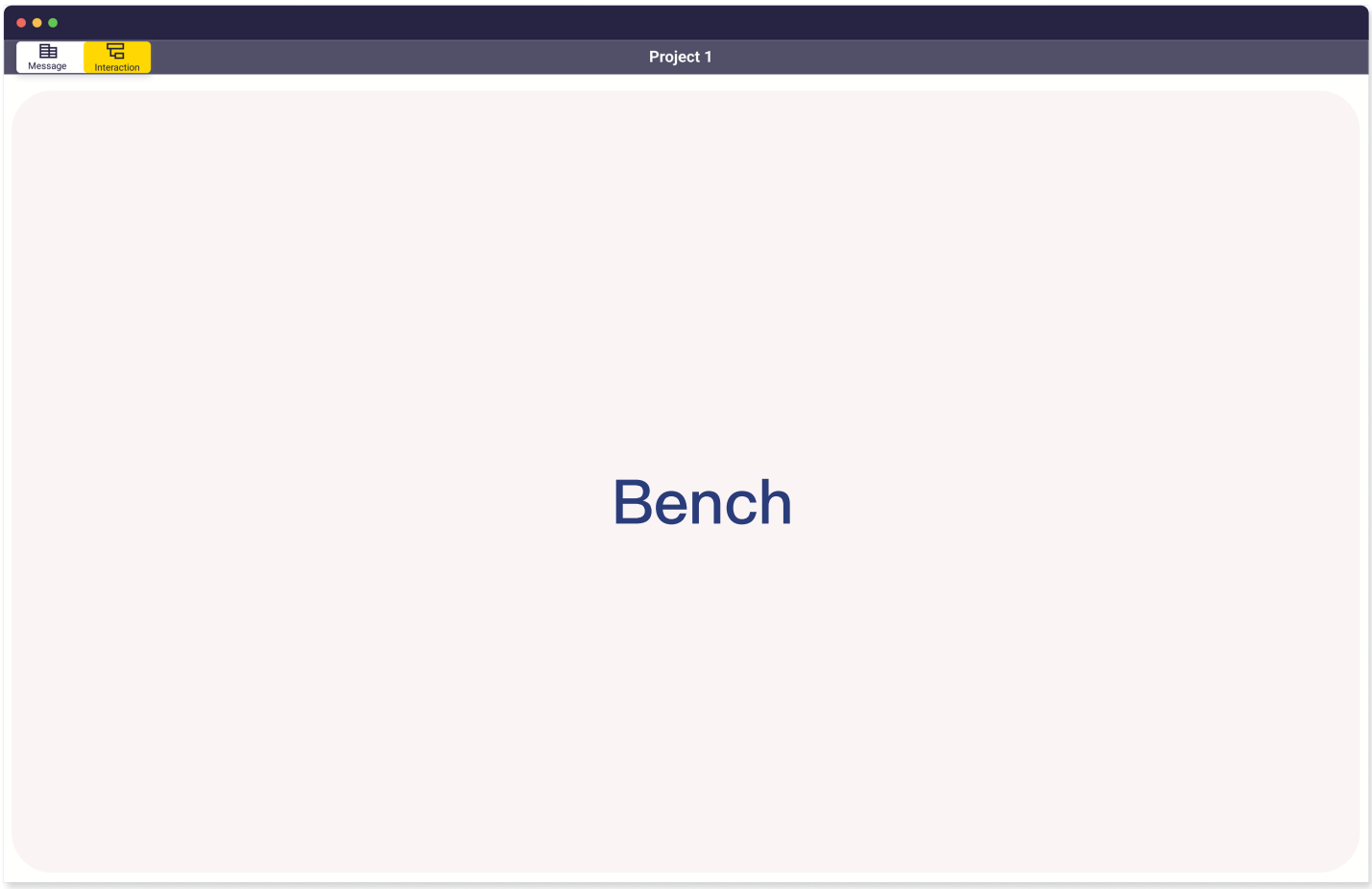


Figure 43.3. Project name.

5. Type of information

In order to reduce the cognitive load of participants in the types of information section and to avoid misinterpretation of the content, "examples" was replaced by "Possible examples" in the new prototype, and the Operation–related content was deleted. Operation–related content was deleted. And because it was difficult for participants to see the yellow button in the upper right corner, it was changed to a darker button in this version, which is visually easier to distinguish.

As shown in the *Figure 43.4*.

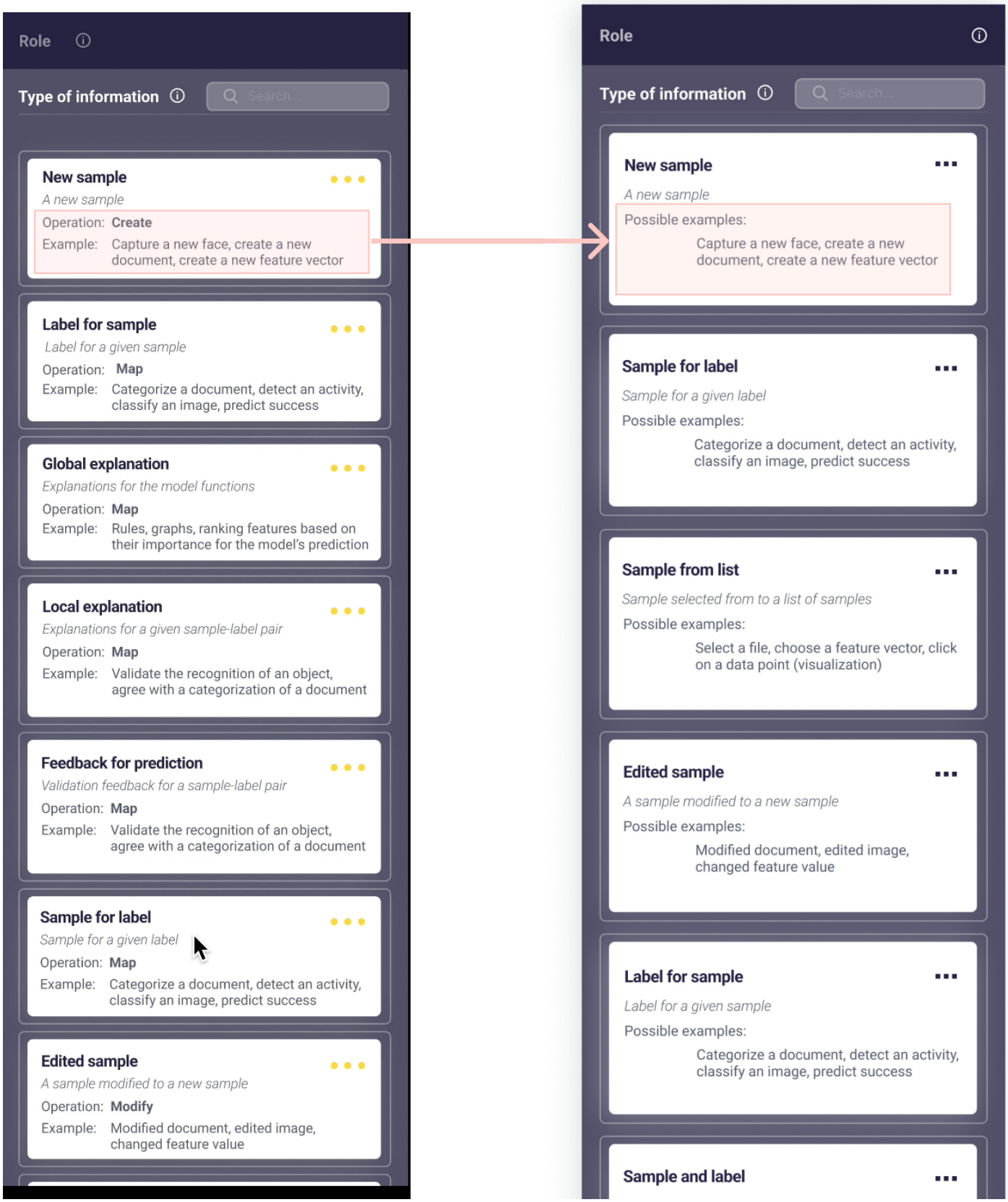


Figure 43.4. Changes on Type of information.

6.1.3 Roles

For the Roles, there is no change on the style of blank Human or Model role part. However, the interaction on how to choose the needed role from the right function part has changed. See *Figure 44*. By clicking the blank Sender or Receiver, the dark-colored function part appears on the right. By draging & dropping the needed role into the responsible blank, it would be fiiled.

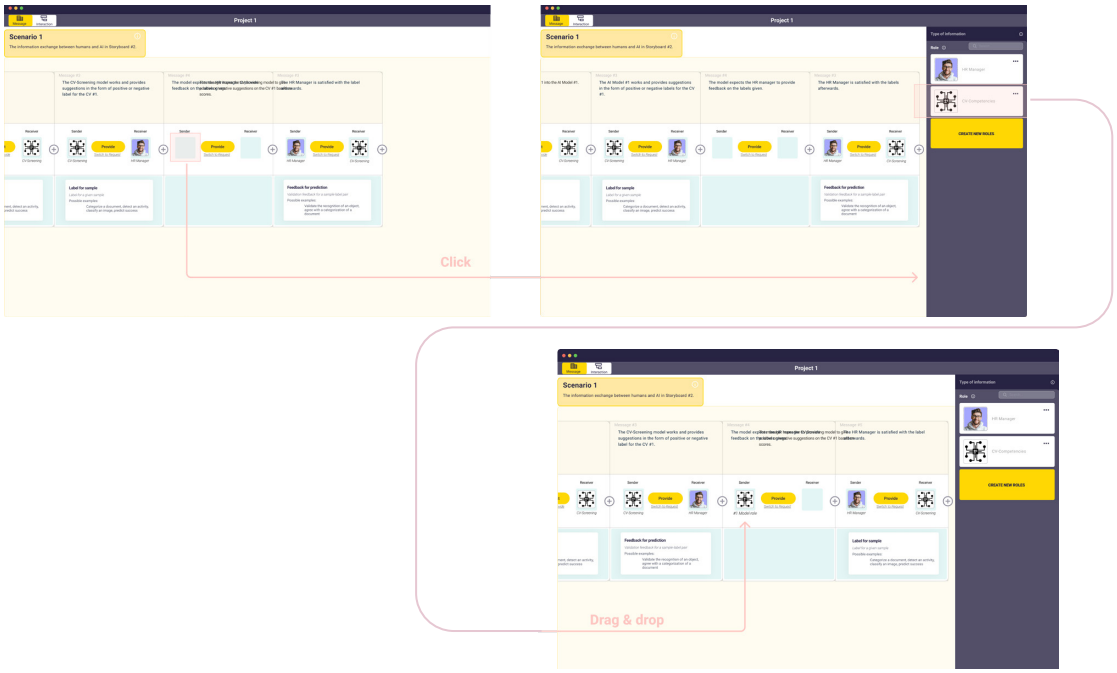


Figure 44. How to fill in blanks for Roles.

6.1.4 Types of information

In the last section, one of the participants pointed out the possibility of needs to see details for two types of information at the same time. Taking this feedback into account, the explanation content for types of information has been modified in the current version to enable the meanwhile viewing of different explanations. See *Figure 45.1*.

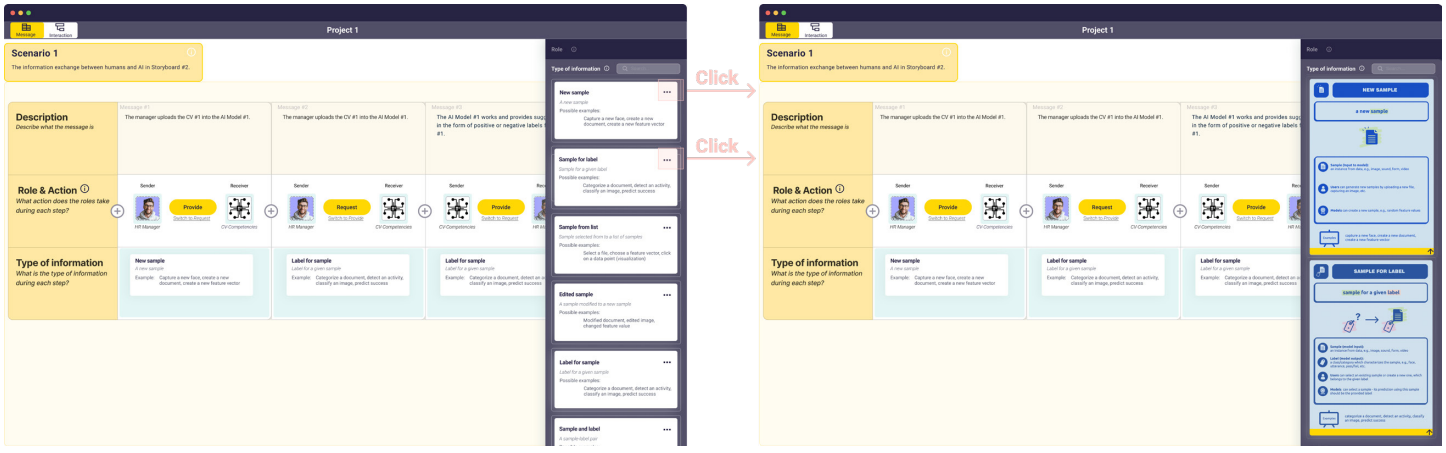


Figure 45.1. See multiple details for different Types of information at the same time..

6.1.5 Touchpoint

In the prior design in Chapter 5, the term "Interaction Row" was what is now named "Touchpoint." Participants argued that Input/Output/UI elements were not in the same hierarchy, making it less fitting to categorize them under a broad term like "Interaction Row." Nevertheless, "Touchpoint," a phrase commonly used in user journey maps, effectively summarize the essence of this section [46]. This term helps designers in understanding the content of this row.

Furthermore, the specifics of Touchpoint have been developed to provide greater precision compared to the earlier version. This redesign aims to offer a more intuitive grasp of what might be present in the interface. See *Figure 45.2 & 45.3*.

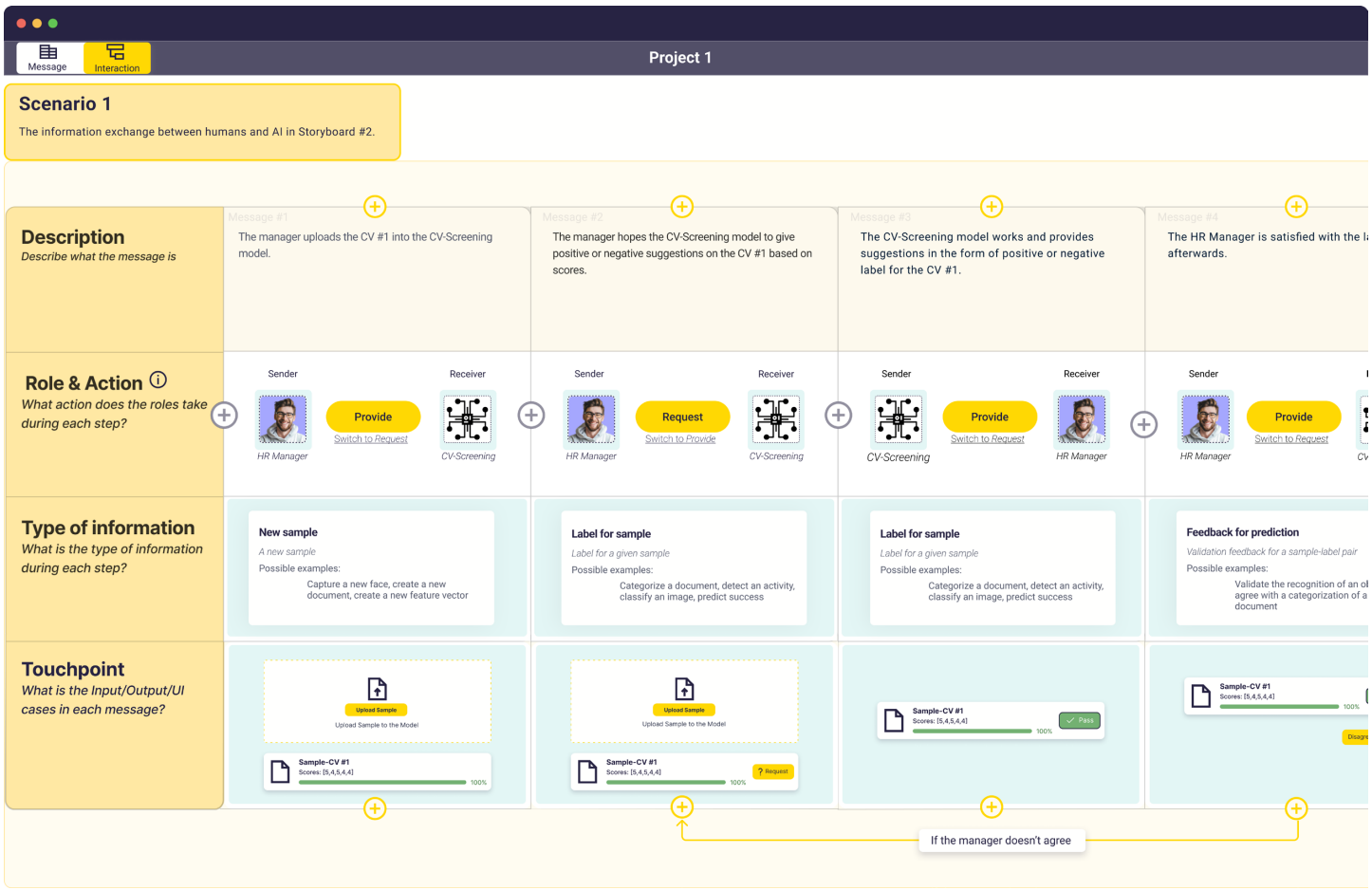


Figure 45.2. Touchpoint and new content.

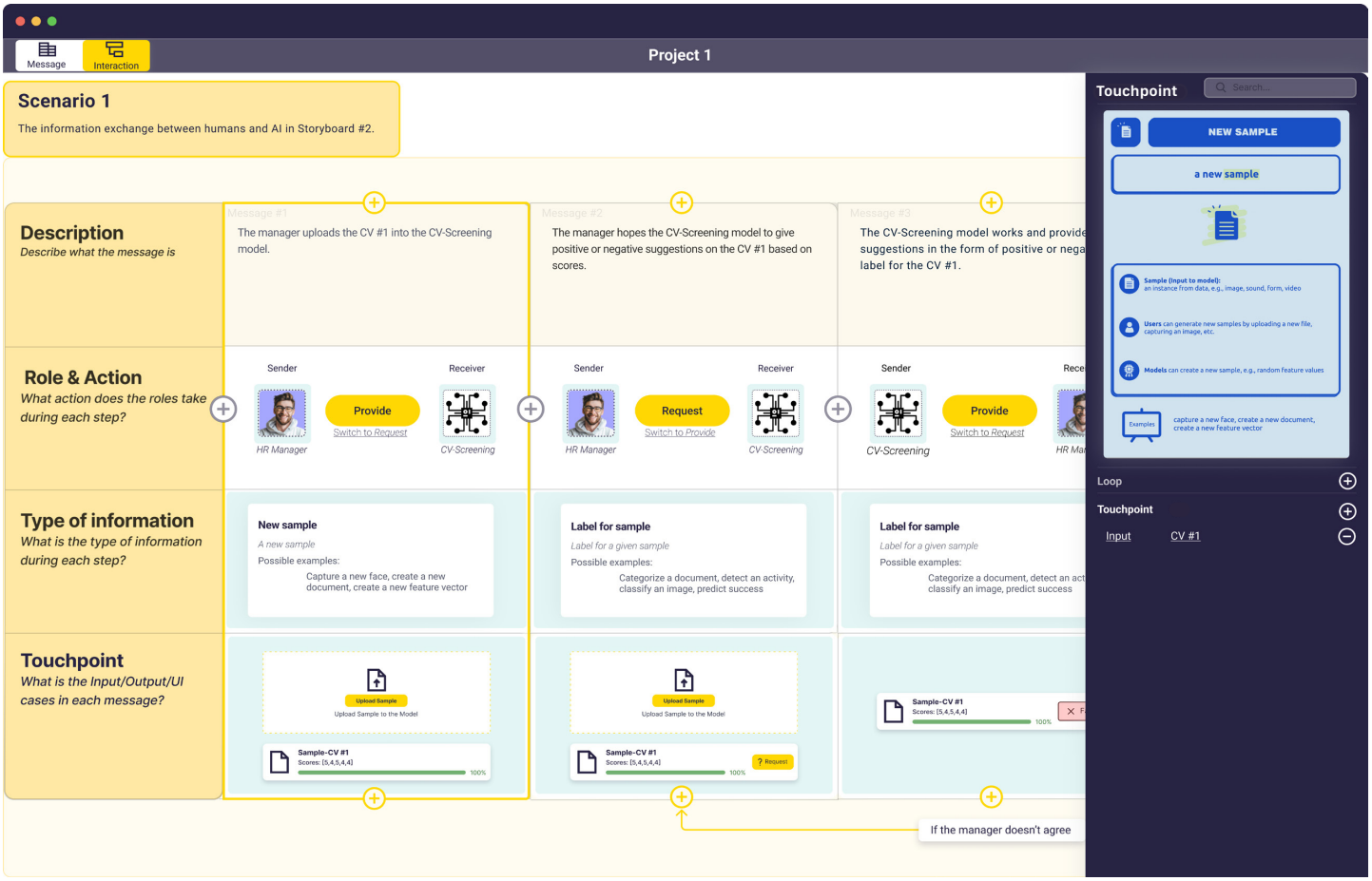


Figure 45.3. Touchpoint and new content – right function part.

6.2 Final test

The test goals and test methods in this section align with the details provided in Chapter 5. The prototype for testing has been enhanced in accordance with the final design.

To improve the testing process, insights gained from Chapter 5 have led to specific refinements. More detailed test reading materials have been added to assist participants in better understanding the use cases applied in the test. This enhancement aims to facilitate participants' interaction with the prototype and successfully complete the tasks. For more details in Appendix J.

6.2.1 Preparation

The Insights part in Chapter 5 has discussed changes. For more details, see them on Appendix J.

6.2.2 During the test

There were six participants in this phase of testing. Two of them were from the SPD program in TUD's IDE Academy and the remaining four were from the DFI program in TUD's IDE Academy. These participants were all new to the program and had no prior knowledge of the terms "communicative acts" involved in the program. Also, none of the five participants had any design experience in creating Human–AI Interactions prior to the test. See *Figure 46*.

For details about the participants, see Appendix K.

For the prototype used in the user test, please click the link: Final prototype

Figure 46. Participants are doing tests.

6.2.3 Analysis & results

Consider the design goals, the responding testable targets and the tasks in Chapter 5.

> *My design goal is to design a digital tool that allows designers to prototype human AI interaction based on communicative acts through a Human-center design process. What's more, this tool:*
>
> *– is easy to learn how to use the tool*
> *– Help designers learn the basic concept "Communicative acts" of human-computer interaction involved in this project.*

Figure 47.1 shows the fulfillment of the testable objectives set in this test. It shows that the design concept in the current prototype, while there is still room for improvement, generally meets the set design goals.

The evaluation of whether or not the testable objectives were met comes from the participants' completion of the individual tasks in *Figure 47.2*. These situations were determined by the participant's behavior, what participants said during the "Think aloud" and the participant's own evaluation during the interview.

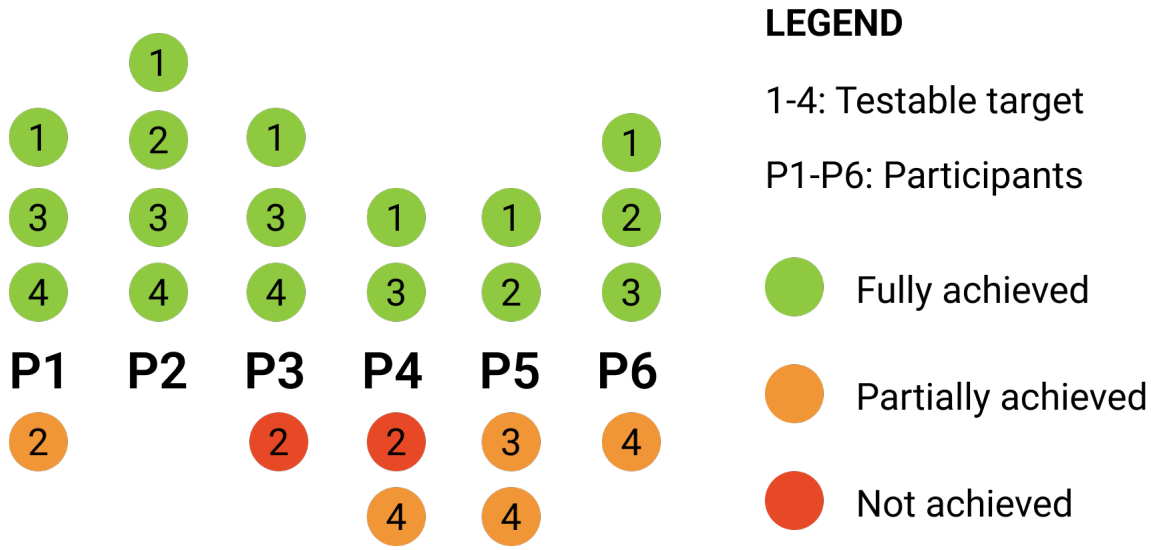


Figure 47.1. Testable targets results.

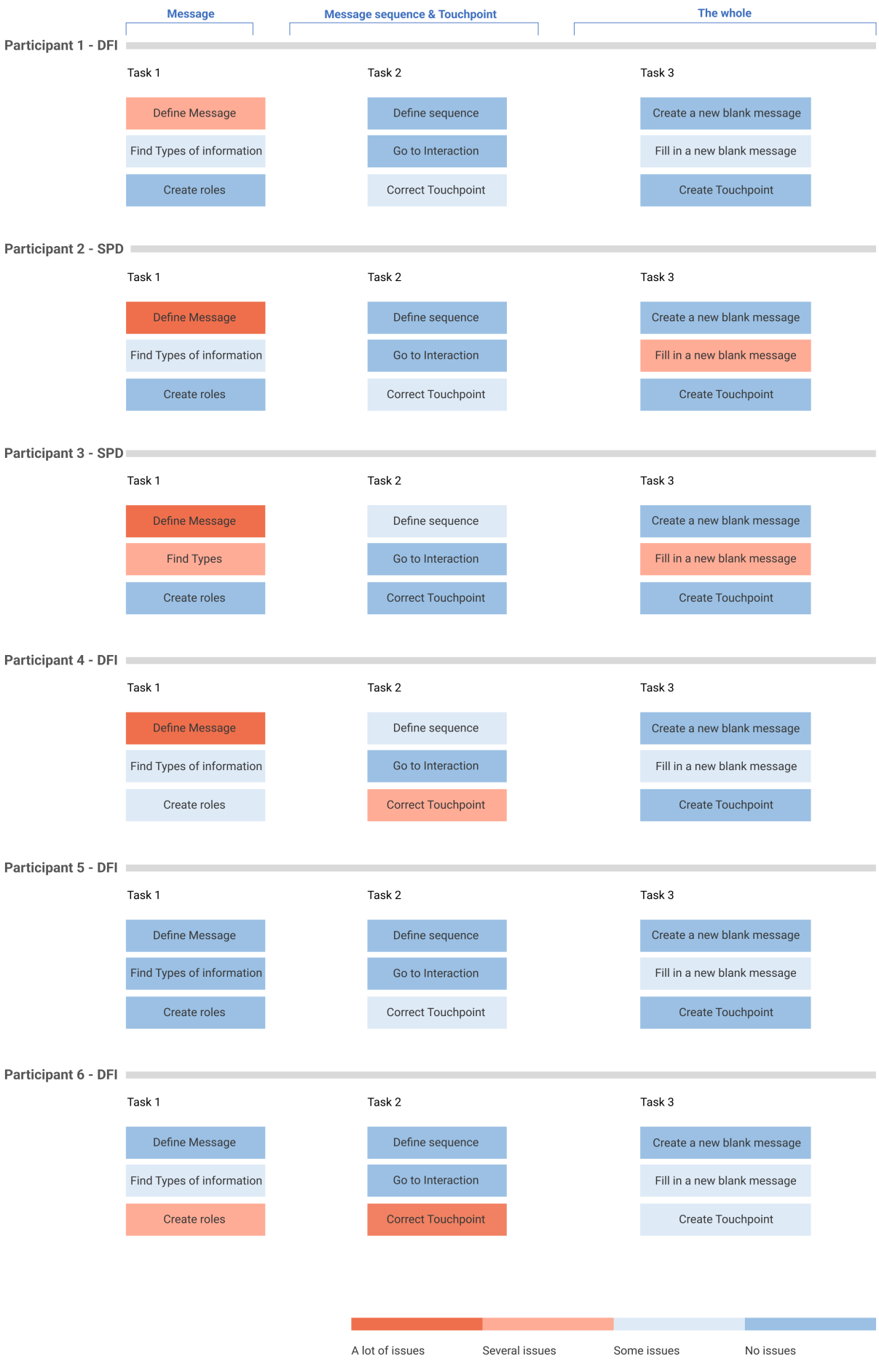


Figure 47.2. Overview of how participants did.

Overall, positive results were obtained for this round of testing; each participant completed three tasks successfully. There are many possible explanations for why participants thought it achieved the design goals from the visulization in *Figure 47.2*.

1. Familiarity helps designers more easily understand the concepts in the project.
First, the way the Message is currently used and presented is easy for participants to associate with the user journey map. this may help them to interpret the Message in the same way they understand the user journey map, and thus quickly understanding the content related to the Communicative acts. See how participants completed Task 1 & 2 in *Figure 47.2*. More than 4 participants said something about it during the test.

"And I get why you asked me what kind of design tools I would use during designing. This looks like a user journey map, and perhaps my familiarity with this tool will also help me subconsciously understand quickly what it takes to make sense of this." —Participant 1
"The prototype has a lot of abstract concepts. It is hard to educate designers how to use it. Probably because I have extensive experience using the user journey map, I could quickly understand how to use this archetype in the first two tasks." —Participant 2
"By the way, I am curious if it has connection with the user journey map. It gives me the same sense." —Participant 5

It is the same for understanding the content in Touchpoint. Touchpoint is one important term in the user journey map and other design tools, which means the ways the target group interact with a product, a brand or a service [45]. It reduces overall understanding and helps participants understand more abstract concepts. There are 4 participants contributing to the insight. See how participants completed "Correct Touchpoint" & "Create Touchpoint" in Figure 47.2.

"It is the information the humans would communicate with the AI model in the use case, right?" —Participant 1
"I think designers would be more familiar with the content in Touchpoint, and that really helps me to understand what happens in each message. Why do you put it on the bottom?" —Participant 4.
"They are the places for human to interact and give triggers for the AI to work." —Participant 5
"Designers would get the touchpoints much easier than the Message." —Participant 6
"The same as the Touchpoint in the user journey map. A classic word in the user journey map." —Participant 6

2. Simplicity helps designers focus more.

The interface is simple and focused, and operates in a way that is consistent with commonly used design tools, balancing the overall difficulty of learning the tool. See *Figure 44.* As shown in the figure, participants only saw Message sequences for the most cases, and only when they needed to make changes did the dark function part appear. This allowed them to immerse themselves in thinking about the relationships between Messages. There are 3 participants talking about the topic directly.

"For the interaction to change the order, I think it is logical. It is the same as other tools." —Participant 1
"I have to say that content about characters, actions and types of information is very important. It would be best to make it more prominent." —Participant 2
"Also, I like designs that hide functional areas as much as possible. This lets me focus more on creating relationships between Messages. This also has a greater view and space." —Participant 4

6.2.4 Insights

The test also exposed many previously unanticipated issues and provided insights for the future development. Mostly it's about the organization of content in Message, the structure of Message sequence and what's in Touchpoint.

1. Organization of content in Messages.

As shown in *Figure 48.1*, the idea among almost all participants was that the current Message format, which combines Description, Roles & Actions, Type of information, and Touchpoint in the same and/or parallel information hierarchy, is not optimal. For them, Roles & Actions and Types of information are the most important sections. While the Touchpoint helped designers in understanding Message and brainstorming potential interactions. In general, the relationship between Roles & Actions and Types of information needs further exploration. The Description section, the same as "annotation," was considered less important and could potentially be hided in the current interface as a lower–level content element.

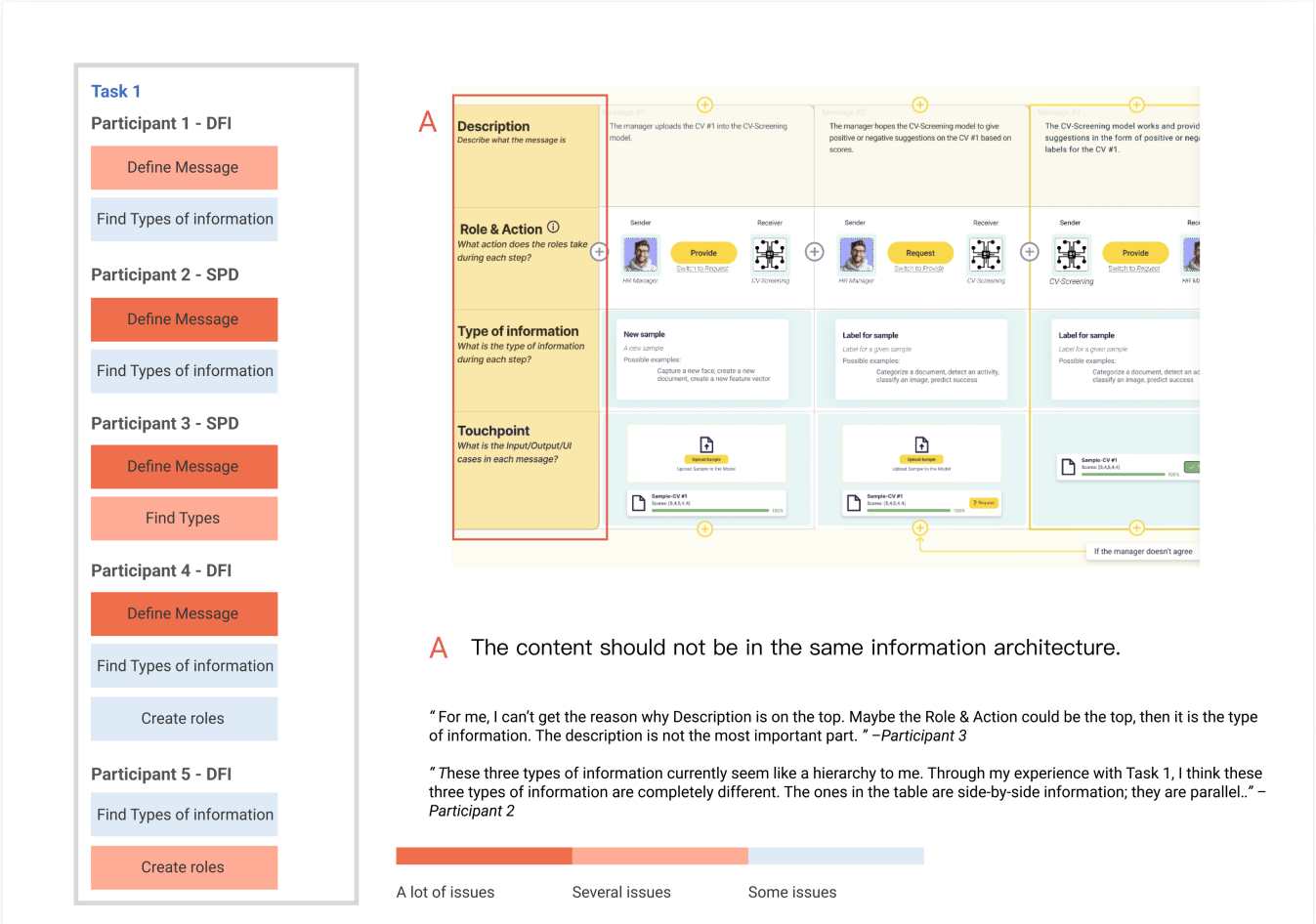


Figure 48.1. Feedback on Message's organization.

2. Explore diverse structures beyond the linear format.

The second part is about the Message sequence's structure. Two participants argued that linear structure is indeed the most efficient structure for information transfer and the most cost–effective way to build a task, it can be challenging for designers in the early stages of ideation. See *Figure 48.2*.

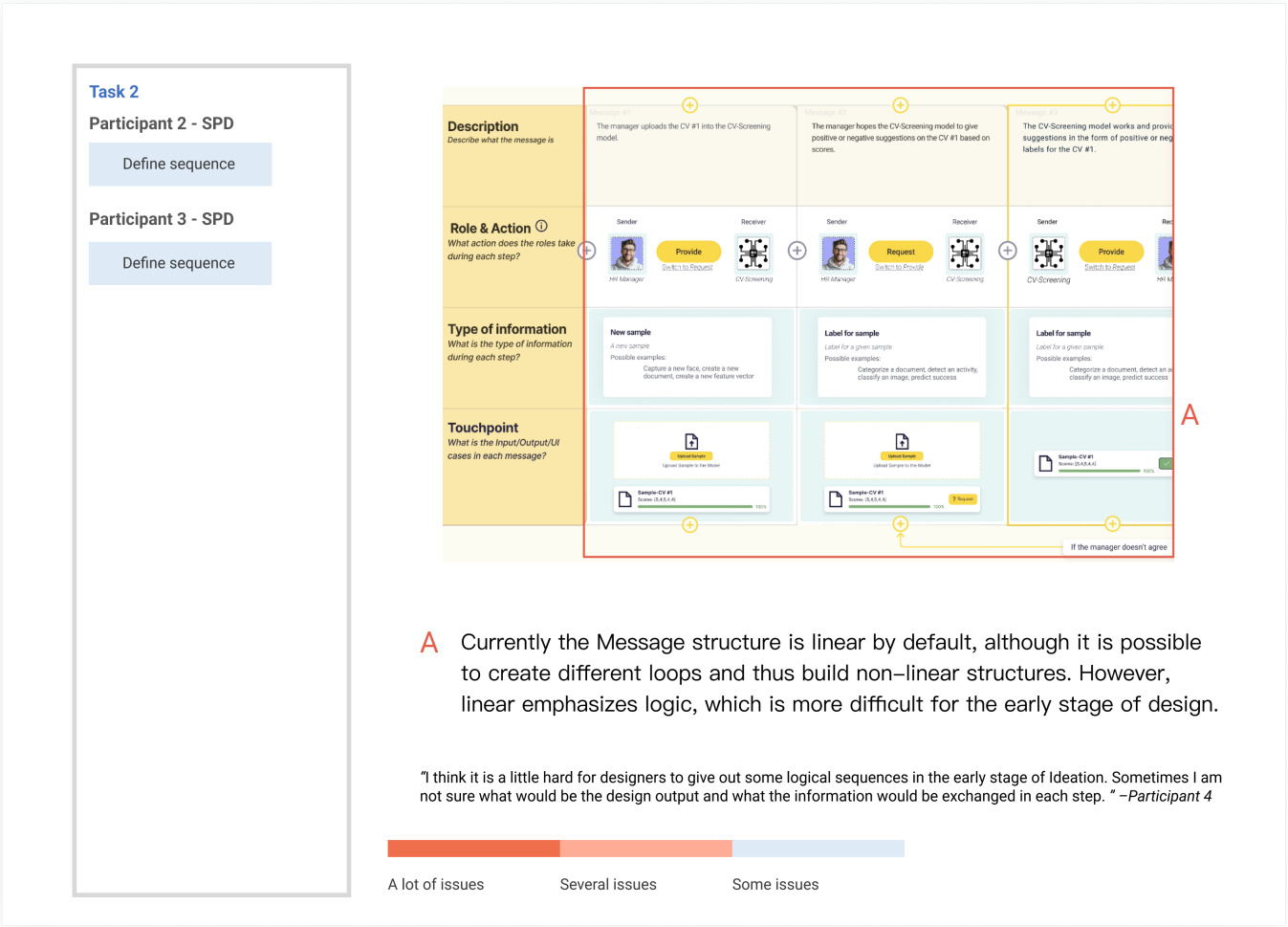


Figure 48.2. Feedback on Message's sequence structure.

As Participant 4 says, *"I think it is a little hard for designers to give out some logical sequences in the early stage of Ideation. am not sure what would be the design output and what the information would be exchanged in each step."*

In general, they prefer to have more control over the design materials at the early stage of the design, so as to explore the different Human–AI Interactions that may exist. As shown in *Figure 48.3*. This is an example presented by Participant 2 of how designers brainstormed in the early stages of the Ideation phase during the Design thinking process in an "everything is in mess" situation, in conjunction with other design tools. They may work in groups to compare and categorize potential design ideas, and then make choices based on actual needs.

This may seem to contradict the conclusions reached in the Pre-phase at the beginning of the project, but this section will be explored in more detail in *6.3 Limitation*.

However, these participants also recognized the benefits of a strong logical structure. It helps them quickly get the content of Human–AI Interactions and construct new information flows. As Participant 4 also said, "*this tool helps me to sort out how HR's needs are to be matched with AI and how his tasks are to be buried at each step.* " Participant 2 mentioned, *"Or keep the linear part, and give them more places to change structures, too."*

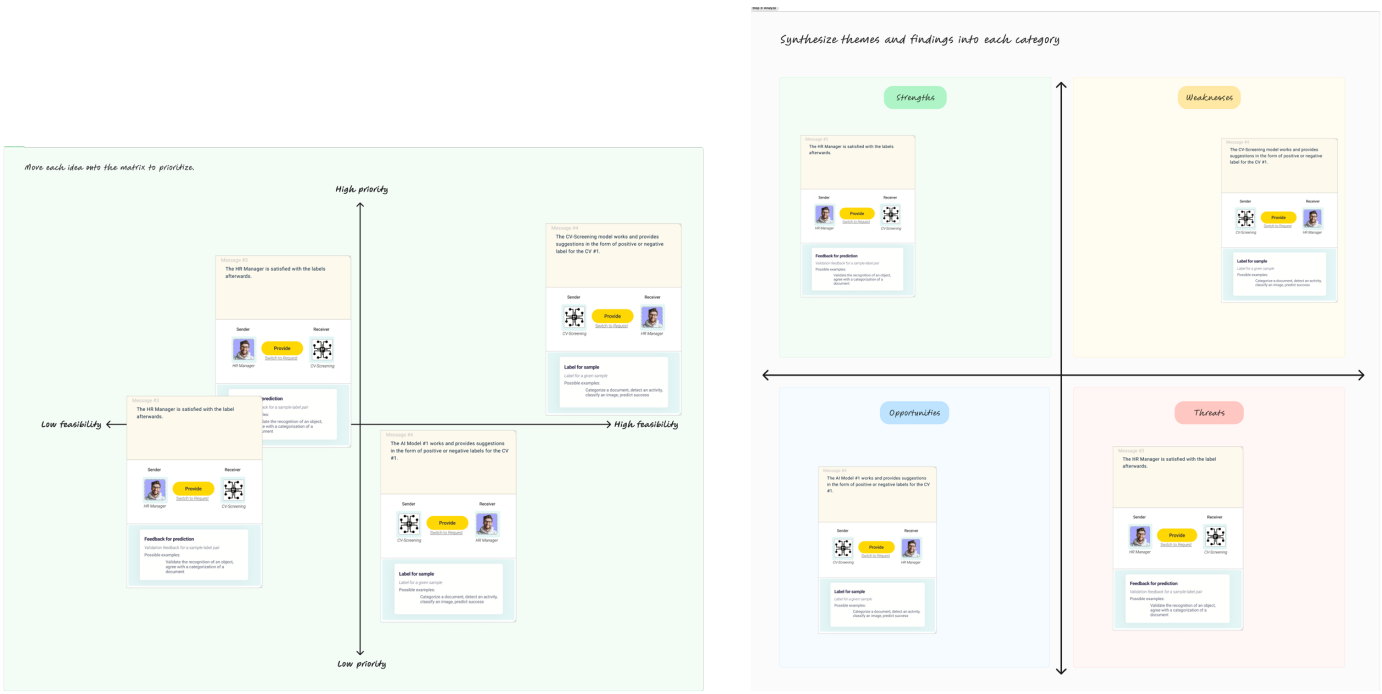


Figure 48.3. Example of one possible way to use Message during ideation.

3. Touchpoint's content has too many details in UI components.

Finally, a suggestion for Touchpoint. As shown in *Figure 48.4*, almost all participants had a problem with the content of Touchpoint. Although they all felt that the current Touchpoint content helps designers to understand the content of Message faster, the current level of detail might hinder designers from generating their own interactive UI and/or interaction ideas later on.

Furthermore, Touchpoint currently lacks the capability to record a designer's fuzzy ideas. Participants believed that it only allowed the selection of prepared UI styles or data, limiting its capacity to inspire designers. Participant 2 highlighted this limitation, stating, *"It is a little strange if I have some ideas but I cannot write or draw it down here, at least for me."*

Of course, there are many other insights in UI styles, interaction and stuff. See more details in Appendix L.

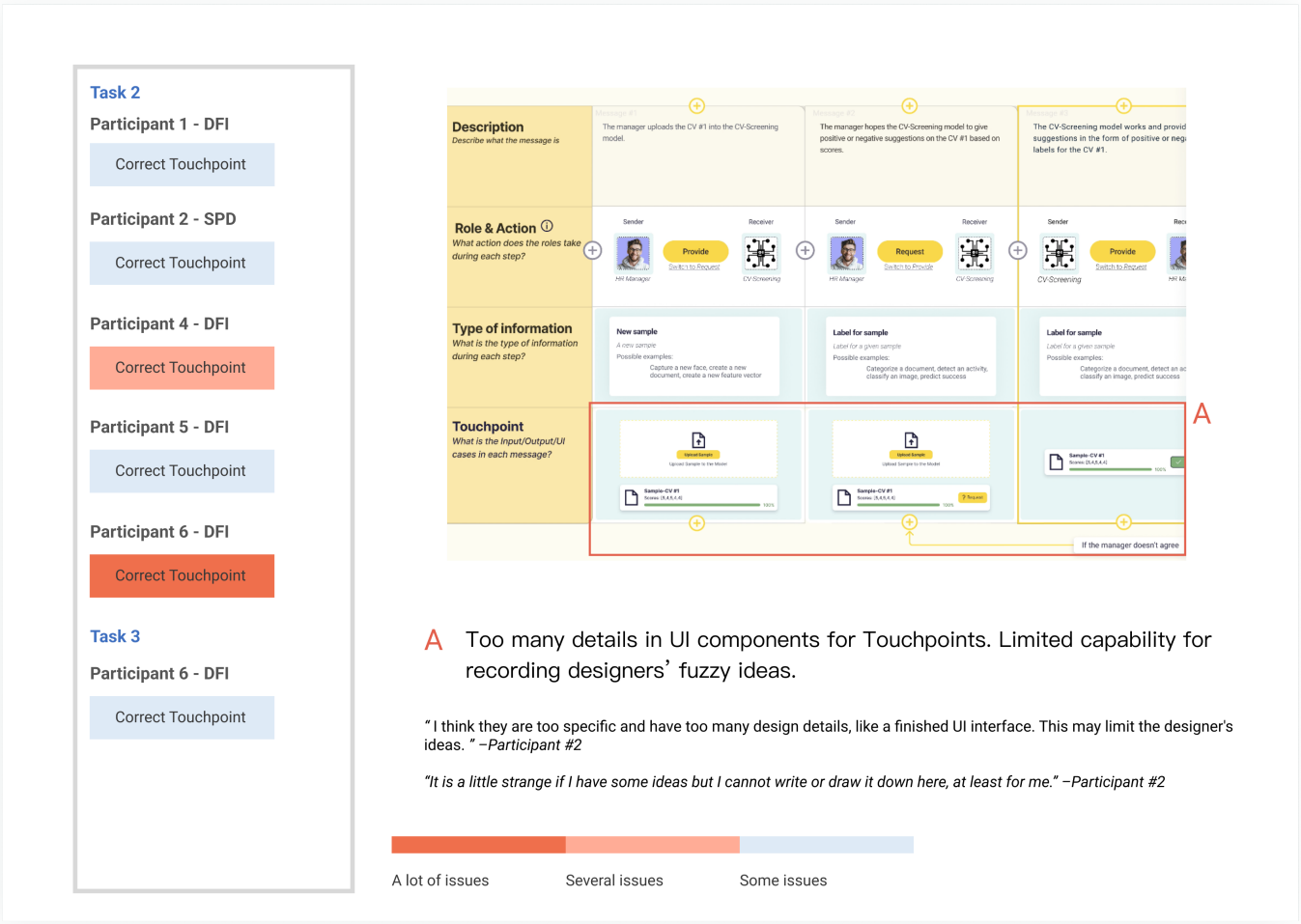


Figure 48.4.Participants feedback on Touchpoint.

6.3 Limitation

The main purpose of this test was to test whether the final design concept met the design goals, so the overall interface was not completed to the point of performing Usability. Therefore, it is unknown if the existing interface can fulfill the requirements related to Usability.

Secondly, this test also uses a use case consistent with the previous test as the context for the overall test task. For more diverse and complex task contexts in the future, whether the current functionality matches the design concepts is also one of the directions that can be tested in the future.

Then from the insights, the tasks setup could be developedin the future. The content of some of the tasks that are set now do not fit well with the activities or behaviors that designers may engage in when they are designing. This has caused participants to become confused about the description or purpose of some tasks, which affects the understanding and use of the prototype. For example in Task 2, Participant 4 means "*But I think it is not realistic. In most cases, the HR would not pay attention to CV's prediction "failed". He or she would just delete it.* "

At the same time, the description of the task needs to ensure that the possibility of misunderstanding is reduced. For example, Participant 3 was confused about the goals of Task 2, which led to a biased understanding of the way the prototype had to work.

"I found Task 2 to be similar to a modification task. May I ask if this is the formalized workflow in the prototype? Is it that after I create the message sequences based on the use case, the prototype automatically generates these Touchpoints for me and I modify them if I am not satisfied?" — Participant 3

Finally, for this type of functional test, the presence of too much information or elements that cannot be interacted with can somewhat interrupt their train of thought. , thus affecting the overall experience of the participants. This is because they may be curious or skeptical about the ways in which these elements are used, and whether the prototype is generating bugs, etc. See *Figure 51*.

"And if in this interface we cannot use the Metric, let it be grey or disappear." —Participant 5.

Similarly, due to the functional limitations of the development tool Figma, the prototypes used for testing were set up in advance with triggers for interactions based on the task, rather than allowing participants to freely explore all interaction features as they saw fit. This also led to an inability to explore in depth how prototypes can be used to create content for vague ideas when the designer is in the ideation phase. This is why the participants' views diverged in Pre–phase and Final phase regarding the test results of the linear structure. The task or activity of the test was on the other hand the cause of this disagreement. In both tasks, a "linear story" had been specified (*From Participant 1 in Pre–test, Pre phase*), so it was easy for the designers to use the linear structure to build a task flow with strong logic. But this is not necessarily the case when they are in the real design process.

Therefore, future research can explore this through a more realistic case with a model with a higher degree of completion.

6.4 Conclusion

In summary, existing digital prototypes can basically achieve design goals through familiarity and simplicity of workflow and interfaces, but the results of the test also show the direction of future design development. The summary about insights in this chapter is shown in *Figure 49.1* below.

Good aspects of the design concept #1	Familiarity helps designers more easily understand the concepts in the project
Good aspects of the design concept #2	Simplicity helps designers focus more
Suggestions for what could be improved #1	Organization of content in Messages
Suggestions for what could be improved #2	Explore diverse structures beyond the linear format
Suggestions for what could be improved #3	Touchpoint's content has too many details in UI components

Figure 49.1. Overview of insights for the final test.

Also, in order to more visually demonstrate how this project developed the research through the insights gained from the different phases of testing, *Figures 49.2 & 49.3 & 49.4* below summarize the insights from Chapters 3 & 5.

And the *Figure 15* shows the specific design considerations from the Pre-test, which combined the literature guidance and the client's supplement.

Suggestions for Design #1	Insights on structures – The linear structure is preferred
Suggestions for Design #2	Insights on Design considerations
Suggestions for Design #3	Communicative acts make sense for participants

Figure 49.2. Overview of insights for the Pre-test.

Design Considerations for model-informed prototyping [13]	Design Considerations for the project	Possible functions
Prototyping tools should allow designers to invoke ML models by specifying input data directly.	1. Designers can create, modify, delete and move messages to form interactions. 2. They can modify the instances or data information about inputs/outputs/instances, etc. of the AI models according to the needs of the end-users for evaluating the predictions.	Create & modify messages & message sequences; choose the best AI models based on the end-users' needs; connect instances in the use case with messages; visualize the input data during the HAI…
Prototyping tools should allow designers to incorporate AI outputs into interface design.	Designers shall be able to visualize in the interface the instances of each Message that are related to the exchange of information or the HAI-related UI elements.	Visualize the output & feedback/XAI; connect instances in the use case with messages; choose the best AI models based on the end-users' needs…
Prototyping tools should allow designers to shape model APIs according to end-user needs.	Designers can define the inputs/outputs/Feedback-XAI of the AI model and the presentation of these based on the needs of the end user in the use case. However, they should not involve too much coding and focus more on how the designer builds the HAI.	Choose the best AI models based on the end-users' needs; help designers understand functions of each AI model easily…
Prototyping tools should allow designers to evaluate design choices across diverse users and contexts.	1. Designers can create personas that will perform HAI based on use cases, whether they are humans or AI models. 2. The design outputs should be broadly applicable to different design challenges.	Create & modify messages & message sequences; connect the end-users' needs with the AI models and messages…
Prototyping tools should allow designers to incorporate model-related data rapidly and iteratively.	Designers should have the flexibility to adapt the content created in the design output to the needs and feedback of the end user.	Create & modify messages & message sequences; visualize the data or instances conveying during the message sequence…

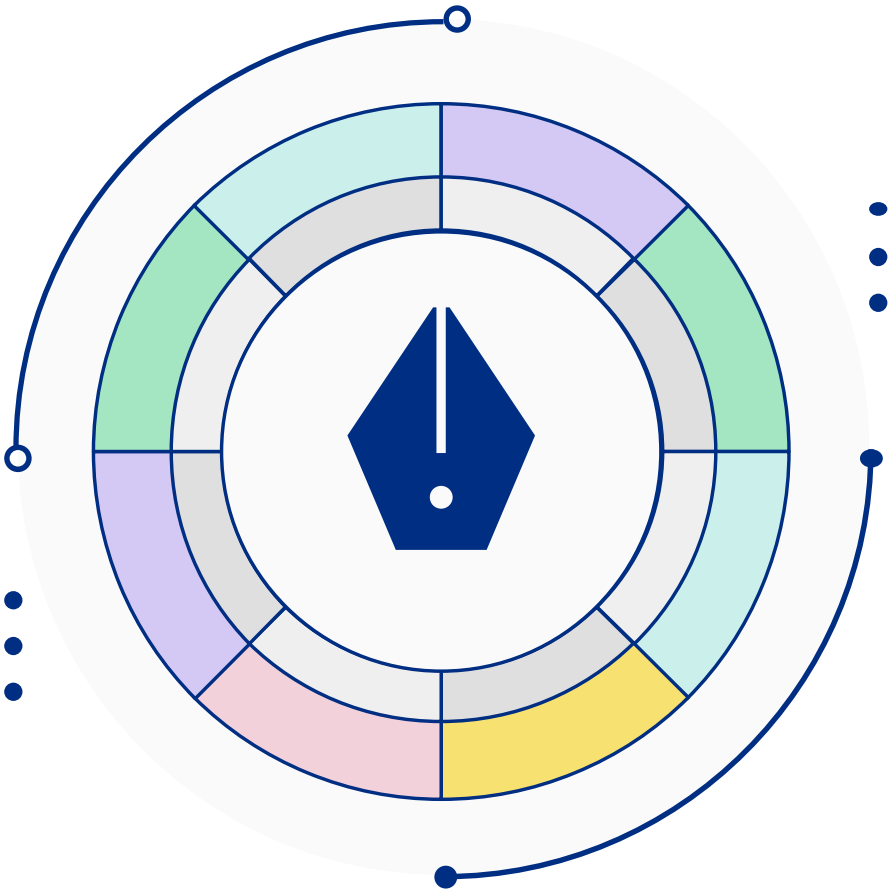
Figure 15. The project's design considerations.

Suggestions for Design #1	Emphasize and clearly demonstrate the connection between Instance and Terms.
Suggestions for Design #2	Fewer interfaces would be better.
Suggestions for Design #3	Be careful to clearly distinguish and present content between different levels in the prototype.
Suggestions for Test #1	Providing a more reasonable test context and supporting materials in the subsequent tests.

Figure 49.3. Overview of insights for the Test 1.

General insight in the design & test plan	The digital prototype and the test plan worked in general
Suggestions for Test #1	The analysis of two pilot tests highlights the necessity for a clearer introduction
Suggestions for Test #2	Display all the parts that can be interacted with where they can be seen most easily, while ensuring that interactions do not interfere with task understanding
Suggestions for Test #3	Reorganize the questions in each task so that they encourage participants to interact with the prototype
Suggestions for Design #1	Change some descriptions in the prototype
Suggestions for Design #2	Keep interactions for the same purpose the same and logical
Suggestions for Design #3	Suggestions on UI components for the prototype

Figure 49.4. Overview of insights for the Test 2 (Pilot test).



7.

Limitation

This section focuses on reflecting on the entire project, including the methodology used for the project, test materials, prototype creation, etc.

With regard to limitations in this whole project, there may be several:

1. ***The linear storytelling approach and task setup may have limited the exploration of more structures.***

From the test results in Pre-phase as well as in Final phase, the use cases CV-Screening used for testing were presented in a linearly descriptive way. Also the strong purposive task setup may have been one of the reasons why participants preferred to use a linear structure in the Pre-phase.

Participants generally acknowledged that linearity was the result of the final HAI presentation because it was strongly logical and more efficient in transferring information. However, in the pre-design phase, they needed more space to find out how the human and the AI would communicate with each other, and to find the opportunity to create the HAI out of "chaos". 2.

2. The interaction constraints of the prototype may have limited the designer's freedom to present ideas.

Due to the limitations of the prototype creation tool Figma does not provide a tool that can really be at the designer's disposal. This allowed participants to accomplish tasks more through the preconceptions. This is not the way designers work in real contexts, so how designers create and conceptualize HAIs in the Ideation phase cannot be captured in finer detail.

3. Message-related information architecture may need to be explored more.

In the project, it is believed that the content in the communicative acts all had equal importance, and subsequent designs were made based on this. However, in the Final phase, it was clear that most participants believed that there was also a hierarchical relationship between these content. If the content could be presented as concisely as possible, it would give designers a clearer picture of how people and AI communicate.

Meanwhile, the "Touchpoint" content in the final design is an important piece of information for designers to understand and utilize communicative acts. This helps designers to transition from concrete concepts to abstract concepts, which is information they can more easily obtain from concrete use cases.

8. Reflection

The topic of "how new technologies can bring better experiences to the lives of the public" has been a concern of mine since I chose to major in design. I believe that it is one of the social responsibilities of designers to deconstruct new technologies and bring the technological breakthroughs in academia to the public to improve their living standards.

And the development of AI over the years and the technological breakthroughs in recent years have made me interested in it. But AI technology is very difficult to understand the content for designers without computer-related background, in my opinion. Meanwhile, during the master programme, although I have been exposed to some AI-related design tools (e.g.Voiceflow, PostureNet, etc.) and use cases of machine learning in design, I still feel that there is always something missing.

Thank you so much to my superviors for giving me the opportunity to explore the possibilities of helping designers build Human-AI Interactions. Your detailed guidance in design, research, and prototyping is greatly appreciated. Although it took a long time to define and align the information for the overall project, we succeeded finally!

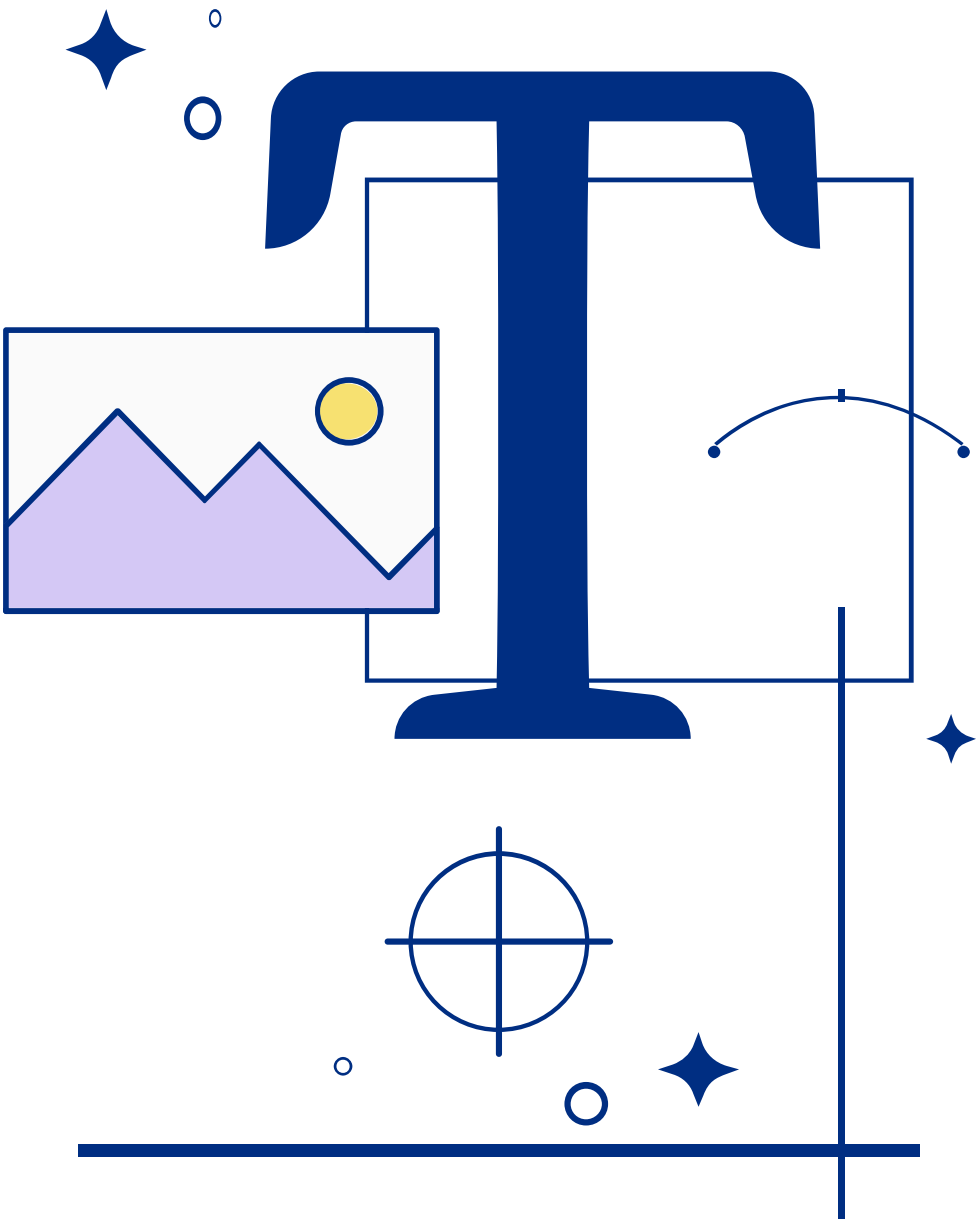
And many thanks to Chatgpt, technology changes lives! It has given a lot of suggestions on the grammar, writing structures and spelling correction for proofreading this thesis.

Although at the beginning of the project I set an ambitious goal to program a truly interactive website, and was initially very confident in my html/css programming skills. Unfortunately, my understanding and use of Javascript was not deep enough, and when it came to more complex interactions, the code always had all kinds of unforeseen problems, and I eventually compromised by using Figma to create interactive prototypes for functional testing. However, this project helped me to review my knowledge of html/css again, and made me realize the gap between the rendering of the design and the final implementation, and how to possibly solve these problems. Maybe this experience can help me communicate more effectively with front-end development engineers in my future job.

Thanks also to Freepik, the final slides and the embellished patterns of this thesis are partly from Freepik [39].

Last but not least, I would like to thank all the participants who took part in the testing of the project, without your support it would have been an impossible task. Also a big thank you to my family members and my cat, not only during this graduation period.

Reference



Reference:

[1] Russell, S. (2022). If We Succeed. *Daedalus*, *151*(2), 43—57. https://doi.org/10.1162/daed_a_01899.

[2] Pavliscak, P. (2016). Algorithms as the new material of design. Retrieved September 14, 2016, from http://www.uxmatters.com/mt/archives/2016/06/algorithms-as-the-new-material-of-design.php.

[3] Spaulding, A., Gajos, K. Z., Jameson, A., Kristensson, P. O., Bunt, A., & Haines, W. (2009). Usable intelligent interactive systems: CHI 2009 special interest group meeting. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems* (pp. 2743-2746).

[4] Wang, Q., Madaio, M., Kane, S., Kapania, S., Terry, M., & Wilcox, L. (2023, April). Designing Responsible AI: Adaptations of UX Practice to Meet Responsible AI Challenges. In *Proceedings of the 2023* Holmquist, L. E. (2017). *Intelligence on tap: artificial intelligence as a new*

design material. interactions, 24 (4), 28-33.

CHI Conference on Human Factors in Computing Systems (pp. 1-16).

[5] Dove, G., Halskov, K., Forlizzi, J., & Zimmerman, J. (2017, May). UX design innovation: Challenges for working with machine learning as a design material. In *Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 278-288).

[6] Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019, May). Improving fairness in machine learning systems: What do industry practitioners need?. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-16).

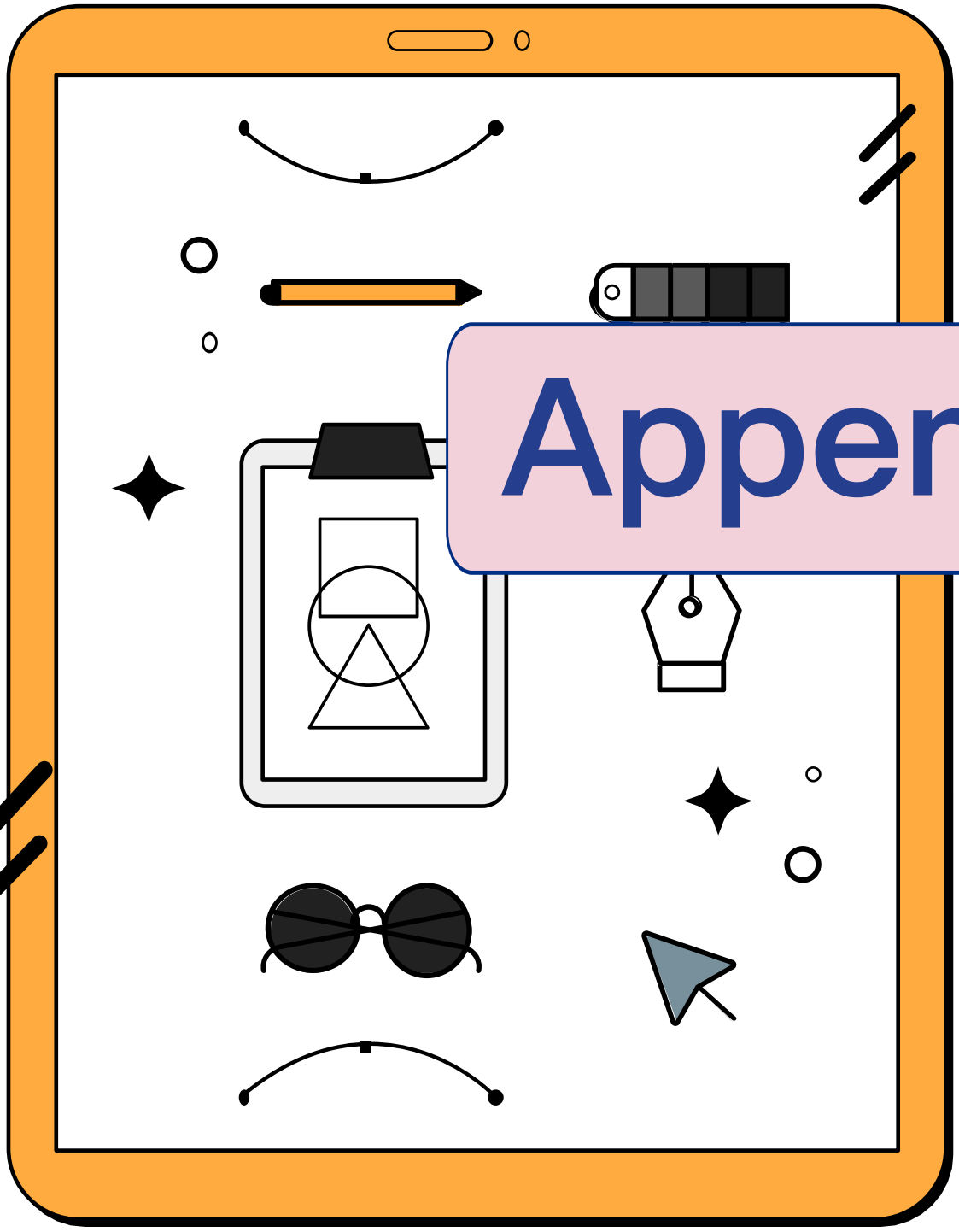
[7] Subramonyam, H., Seifert, C., & Adar, E. (2021, June). Towards a process model for co-creating AI experiences. In *Designing Interactive Systems Conference 2021* (pp. 1529-1543).

[8] Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020, April). Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1-13).

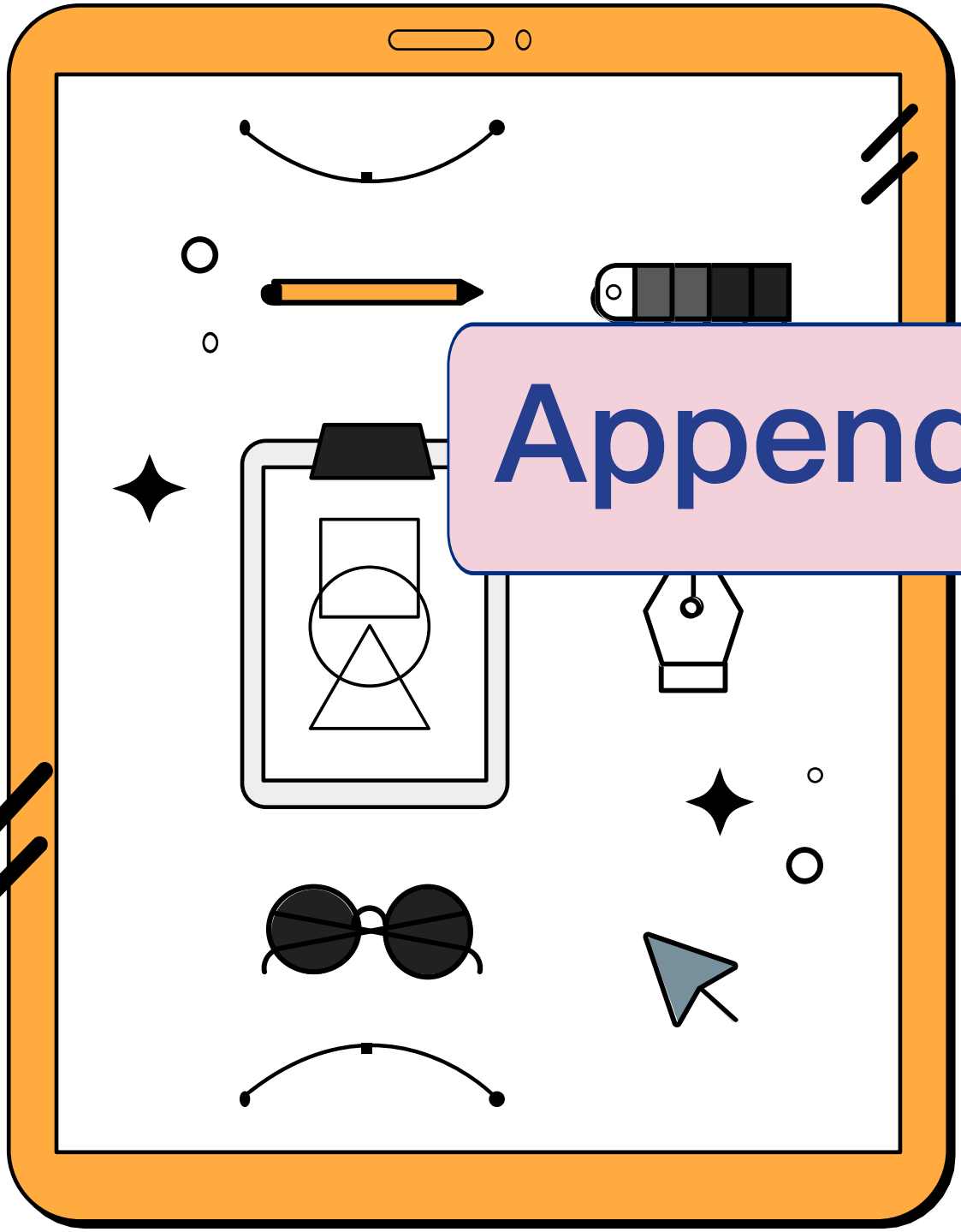
[9] Unpacking Human–AI Interactions: From Interaction Primitives to a Design Space, Tsiakas, K and Murray–Rust, D. In ACM Transactions on Interactive Intelligent Systems (just accepted).
[10] Goldman, N., & Narayanaswamy, K. (1992, June). Software evolution through iterative prototyping. In *Proceedings of the 14th international conference on Software engineering* (pp. 158–172).
[11] Pacific Research Laboratories. (n.d.). 5 Advantages of Iterative Design and Prototyping. Retrieved December 4, 2023, from https://www.pacific–research.com/5–advantages–of–iterative–design–and–prototyping–prl/
[12] Google. 2019. People + AI Guidebook. https://pair.withgoogle.com/
[13] Apple Inc. 2019. Designing the UI and User Experience of a Machine Learning App. https://developer.apple.com/design/human–interface–guidelines/machinelearning/overview/introduction/
[14] Subramonyam, H., Seifert, C., & Adar, E. (2021, April). Protoai: Model–informed prototyping for ai–powered interfaces. In *26th International Conference on Intelligent User Interfaces* (pp. 48–58).
[15] Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development, 3*(3), 210–22
[16] Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., ... & Teller, A. (2022). Artificial intelligence and life in 2030: the one hundred year study on artificial intelligence. *arXiv preprint arXiv*:2211.06318 .
[17] Yang, Q., Banovic, N., & Zimmerman, J. (2018, April). Mapping machine learning advances from hci research to reveal starting places for design innovation. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–11)
[18] Klemmer, S. R., Sinha, A. K., Chen, J., Landay, J. A., Aboobaker, N., & Wang, A. (2000, November). Suede: a wizard of oz prototyping tool for speech user interfaces. In *Proceedings of the 13th annual ACM symposium on User interface software and technology* (pp. 1–10).
[19] Dix, A. (2003). Human–computer interaction. Pearson Education.
[20] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human–AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–13).

[21] Maadi, M., Akbarzadeh Khorshidi, H., & Aickelin, U. (2021). A review on human—ai interaction in machine learning and insights for medical applications. *International journal of environmental research and public health, 18*(4), 2121.
[22] OpenAI. (2023). ChatGPT [Large language model]. https://chat.openai.com
[23] Holmquist, L. E. (2017). *Intelligence on tap: artificial intelligence as a new design material. interactions, 24* (4), 28–33.
[24] Cartwright, A. (2016). Invisible Design: Co–Designing with Machines.
[25] Yang, Q., Zimmerman, J., Steinfeld, A., & Tomasic, A. (2016, June). Planning adaptive mobile experiences when wireframing. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (pp. 565–576).
[26] Cranshaw, J., Elwany, E., Newman, T., Kocielnik, R., Yu, B., Soni, S., ... & Monroy–Hernández, A. (2017, May). Calendar. help: Designing a workflowbased scheduling agent with humans in the loop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 2382–2393).
[27] Klemmer, S. R., Sinha, A. K., Chen, J., Landay, J. A., Aboobaker, N., & Wang, A. (2000, November). Suede: a wizard of oz prototyping tool for speech user interfaces. In *Proceedings of the 13th annual ACM symposium on User interface software and technology* (pp. 1–10).
[28] Yang, Q., Cranshaw, J., Amershi, S., Iqbal, S. T., & Teevan, J. (2019, May). Sketching nlp: A case study of exploring the right things to design with language intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12).
[29] Girardin, F., & Lathia, N. (2017, March). When user experience designers partner with data scientists. In *2017 AAAI Spring Symposium Series* .
[30] Bly, S., & Churchill, E. F. (1999). Design through matchmaking: technology in search of users. *interactions, 6* (2), 23–31.
[31] Boehner, K., & DiSalvo, C. (2016, May). Data, design and civics: An exploratory study of civic tech. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2970–2981).
[32] Tuttle, G. (2021). What is design thinking and why is it important?. *We Work Ideas, 24*.

[33] Interaction Design Foundation – IxDF. (2016, May 25). *What is Design Thinking?* . Interaction Design Foundation – IxDF. https://www.interaction-design.org/literature/topics/design-thinking
[34] Dekker, R. (2022). Designing AI projects using a reactive digital canvas. http://resolver.tudelft.nl/uuid:105abeee-6430-4eb7-b3fd-eda35add900d
[35] Voiceflow. (n.d.). Retrieved from https://www.voiceflow.com/
[36] AI Meets Design Toolkit — AIxDESIGN. (n.d.). AIxDESIGN. https://aixdesign.co/toolkit
[37] Difference between Linear and Non-linear Data Structures. (2023, February 28th). Retrieved August 1st, 2023, from: https://www.geeksforgeeks.org/difference-between-linear-and-non-linear-data-structures/
[38] Endmann, A. & Keßner, D. (2016). User Journey Mapping — A Method in User Experience Design. *i-com, 15* (1), 105–110. https://doi-org.tudelft.idm.oclc.org/10.1515/icom-2016-0010
[39] 7 Ways to Analyze a Customer-Journey Map. (n.d.). Nielsen Norman Group. https://www.nngroup.com/articles/analyze-customer-journey-map/
[40] Chin, S. Figma profile [Profile]. Retrieved from https://www.figma.com/@seanchin
[41] Riddle UI. Retrieved from https://riddleui.com
[42] Freepik. Illustration with multitasking concept [Digital image]. Retrieved from https://www.freepik.com/free-vector/illustration-with-multitaskingconcept_6555998.htm
[43] Shakya, R. Figma profile [Profile]. Retrieved from https://www.figma.com/@shakya90
[44] UXIS. Figma profile [Profile]. Retrieved from https://www.figma.com/@uxis
[45] Van Someren, M., Barnard, Y. F., & Sandberg, J. (1994). The think aloud method: a practical approach to modelling cognitive. *London: AcademicPress, 11*, 29–41.
[46] Interaction Design Foundation – IxDF. (2023, November 21). *What is Human-AI Interaction?* . Interaction Design Foundation – IxDF. https://www.interaction-design.org/literature/topics/human-ai-interaction
[47] Wikipedia contributors. (2023, November 18). Touchpoint. Wikipedia. https://en.wikipedia.org/wiki/Touchpoint



Appendix



Appendix A

DESIGN FOR our future

TU Delft

IDE Master Graduation

Project team, Procedural checks and personal Project brief

This document contains the agreements made between student and supervisory team about the student's IDE Master Graduation Project. This document can also include the involvement of an external organisation, however, it does not cover any legal employment relationship that the student and the client (might) agree upon. Next to that, this document facilitates the required procedural checks. In this document:

- The student defines the team, what he/she is going to do/deliver and how that will come about.
- SSC E&SA (Shared Service Center, Education & Student Affairs) reports on the student's registration and study progress.
- IDE's Board of Examiners confirms if the student is allowed to start the Graduation Project.

USE ADOBE ACROBAT READER TO OPEN, EDIT AND SAVE THIS DOCUMENT

Download again and reopen in case you tried other software, such as Preview (Mac) or a webbrowser.

STUDENT DATA & MASTER PROGRAMME

Save this form according the format "IDE Master Graduation Project Brief_familyname_firstname_studentnumber_dd-mm-yyyy".
Complete all blue parts of the form and include the approved Project Brief in your Graduation Report as Appendix 1 !

family name: Zhou
initials: J given name: Jiayi
student number: 5568935
street & no.:
zipcode & city:
country:
phone:
email:

Your master programme (only select the options that apply to you):

IDE master(s): ( ) IPD (★) DfI ( ) SPD
2nd non-IDE master:
individual programme: - - (give date of approval)
honours programme: ( ) Honours Programme Master
specialisation / annotation: ( ) Medisign
( ) Tech. in Sustainable Design
( ) Entrepeneurship

SUPERVISORY TEAM **

Fill in the required data for the supervisory team members. Please check the instructions on the right !

** chair: Dave Murray-Rust dept. / section: IDE/HCD
** mentor: Mahan Mehrvarz dept. / section: IDE/HCD
2nd mentor: Konstantinos Tsiakas
organisation: IDE/HCD
city: country:

comments (optional)

Chair should request the IDE Board of Examiners for approval of a non-IDE mentor, including a motivation letter and c.v..

Second mentor only applies in case the assignment is hosted by an external organisation.

Ensure a heterogeneous team. In case you wish to include two team members from the same section, please explain why.

APPROVAL PROJECT BRIEF

To be filled in by the chair of the supervisory team.

chair Dave Murray-Rust date 1 - 8 - 2023 signature

CHECK STUDY PROGRESS

To be filled in by the SSC E&SA (Shared Service Center, Education & Student Affairs), after approval of the project brief by the Chair.
The study progress will be checked for a 2nd time just before the green light meeting.

Master electives no. of EC accumulated in total: 36 EC

Of which, taking the conditional requirements into account, can be part of the exam programme 30 EC

List of electives obtained before the third semester without approval of the BoE

★ YES all 1st year master courses passed

NO missing 1st year master courses are:

name Robin den Braber date 02 - 08 - 2023 signature

Robin den Braber Digitaal ondertekend door Robin den Braber Datum: 2023.08.02 08:37:53 +02'00'

FORMAL APPROVAL GRADUATION PROJECT

To be filled in by the Board of Examiners of IDE TU Delft. Please check the supervisory team and study the parts of the brief marked **.
Next, please assess, (dis)approve and sign this Project Brief, by using the criteria below.

- Does the project fit within the (MSc)-programme of the student (taking into account, if described, the activities done next to the obligatory MSc specific courses)?
- Is the level of the project challenging enough for a MSc IDE graduating student?
- Is the project expected to be doable within 100 working days/20 weeks ?
- Does the composition of the supervisory team comply with the regulations and fit the assignment ?

Content: ★ APPROVED NOT APPROVED

Procedure: ★ APPROVED NOT APPROVED

comments

name Monique von Morgen date 22 - 08 - 2023 signature

Initials & Name J Zhou Student number 5568935

Title of Project Develop an interface for model-informed prototyping of HAI interactions

Develop an interface for model-informed prototyping of HAI interactions

project title

Please state the title of your graduation project (above) and the start date and end date (below). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

start date 24 - 07 - 2023 08 - 12 - 2023 end date

INTRODUCTION **

Please describe, the context of your project, and address the main stakeholders (interests) within this context in a concise yet complete manner. Who are involved, what do they value and how do they currently operate within the given context? What are the main opportunities and limitations you are currently aware of (cultural- and social norms, resources (time, money,...), technology, ...).

Every advancement in technology opens up possibilities for designers to innovate. Often finding the right way to use technology requires designers to make an effort to learn and deconstruct the technology. Artificial Intelligence technology, after close to a century of development, has enough resources and is at the stage of entering the design market [1]. Existing research proves that AI has the potential to become a new material for user experience[2].

However, user interface or user experience designers have found great challenges in investigating how to utilize AI as UX design material. Due to the technical complexity and uncertainty associated with AI models [3], designers face difficulties in effectively designing and testing these interface prototypes. Specifically, these two characteristics result in the fact that human traditional prototyping techniques may not be able to adequately capture the behavior and capabilities of AI systems [3]. Therefore, it is beneficial to investigate how to bridge the gap between design and engineering practices to ensure that AI products produce optimal performance.

As a result, Our main goal is to develop a prototyping tool for AI interactions using communicative behavior (Figure 1). Communicative behavior (also known as Communicative acts) means intentional actions performed by individuals to convey meaning, exchange information, express emotions or achieve specific goals through language or other forms of communication. These acts would be like speaking/writing languages, gestures, facial expressions, body languages and more. They are very important because they are the focus of understanding or transferring information between humans and AI. As a prototyping tool, we expect to be able to efficiently receive information about the user's needs and deliver it to the user in an appropriate way. To collect these communicative behavior from designers, Model-Informed prototyping(MIP) would be a great way [4]. MIP is a workflow that combines model exploration and interface design tasks. That means we could see what kind of communicative behavior happen and how designers interact with AI and interfaces while designers are in the process of MIP.

The tool will be developed using html/css/Javascript and P5.js, allowing designers to overcome the challenges posed by AI and create seamless and effective AI-powered user interfaces.

When developing a prototyping tool for HAI interactions, the primary goal is to maximize the ability to ensure that the product provides correct feedback on end-user behavior. In this regard, good HAI design guidelines have been proposed by both academia and industry [5, 6, 7]. I would conclude design considerations during the design part.

Opportunities I am currently aware of:
- Artificial intelligence materials challenge traditional prototyping methods because it requires more input to create interactive prototypes and give immediate feedback.
- There is need for methodological research on how to let UX experience designers interact with AI materials with feedback, and we can look for possible directions by means of rapid prototyping and so on.
- Friction in prototyping AI features for designers.
- Lower the barrier for data-driven design and more support in current prototyping tools.
- Concluding our approach from the project into results or models which could support the future on a new workflow for HAI prototyping.

space available for images / figures on next page

Initials & Name J Zhou Student number 5568935

Title of Project Develop an interface for model-informed prototyping of HAI interactions

introduction (continued): space for images

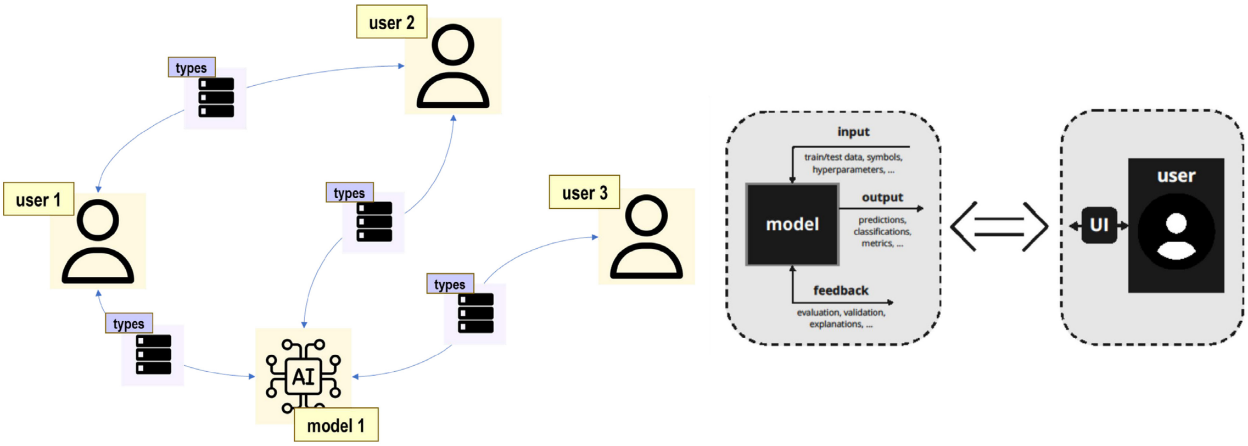


image / figure 1: Interactions and types of communicated information between users and AI models

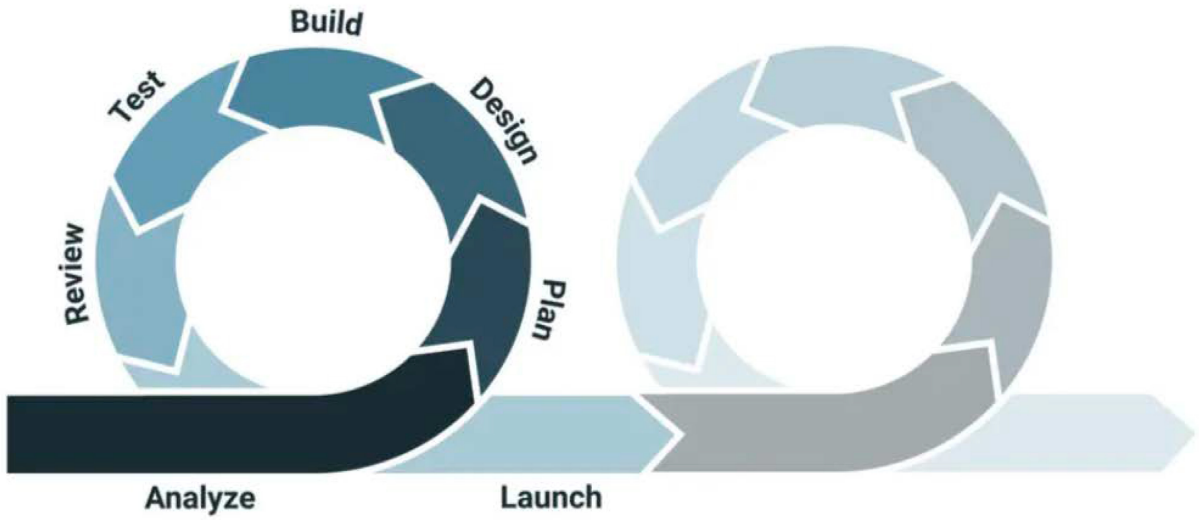


image / figure 2: Iterative protyping process

Initials & Name J Zhou Student number 5568935

Title of Project Develop an interface for model-informed prototyping of HAI interactions

PROBLEM DEFINITION **

Limit and define the scope and solution space of your project to one that is manageable within one Master Graduation Project of 30 EC (= 20 full time weeks or 100 working days) and clearly indicate what issue(s) should be addressed in this project.

The main goal is to develop a tool to prototype HAI interactions using a low-level type of communicative information, communicative behavior (Fig. 1). The entire design and research process will be guided by an iterative prototyping approach [8], where each phase consists of four steps (Fig. 2), for a total of four phases: 1) The pre-prototype phase is a rapid prototyping phase. The main objective is to quickly gain basic insights from design students on prototyping tools such as representing and creating HAI interactions through paper models. Participants may have experience in the use of HAI as a UX material and may bring some unique insights to the next prototyping development. Ideally I could have a draft design considerations or assessment after the workshop. 2) Phase 1 consists of brainstorming, results of the pre-prototyping and idea realization of the first version of the digital prototyping software, using Figma as the initial model building tool. The focus of this phase is on the usability of the prototype, including task flow, interface elements, interaction styles, and so on. The test is expected to recruit 5-10 participants (design students). The test will use Model - Informed Prototyping (MIP) to create a product using the model for the same use case. The prototype will then be iterated based on the results. 3) Phase 2 is the same as Phase 1. However, this phase focuses on whether the improved workflows, interactions, system feedback, etc. During this phase, we would use html/css/javascript to build the real web interfaces. This phase of testing is expected to recruit 15-20 participants and again have participants complete tasks using the MIP. Based on the results of the testing, I will iterate on the prototype again. 4) Phase 3 focuses on testing the final prototype to see if it meets the requirements of different use cases. We would also use the html/css/javascript to build the prototype. The test is tentatively set for 2 different use cases with a total of 15-21 participants recruited.

ASSIGNMENT **

State in 2 or 3 sentences what you are going to research, design, create and / or generate, that will solve (part of) the issue(s) pointed out in "problem definition". Then illustrate this assignment by indicating what kind of solution you expect and / or aim to deliver, for instance: a product, a product-service combination, a strategy illustrated through product or product-service combination ideas, ... In case of a Specialisation and/or Annotation, make sure the assignment reflects this/these.

I will develop an interface for model-informed prototyping using HAI communicative acts. To do it, I would use Iterative Prototyping to explore how designers interact with the interface while prototyping HAI interactions. Then I would assess and evaluate prototypes based on results, including both quantitative and qualitative data.

- 1) Make a plan for pre/1st/2nd/3rd test, including participants, use cases, assessments, questionnaires and so on.
- 2) Brainstorm for the pre/1st/2nd/3rd prototypes. The deliverable an outcomes could be paper models or clickable digital prototypes which participants could use for MIP.
- 3) Build pre/1st prototypes by paper or Figma materials. Build the 2nd/3rd prototype by html/CSS/Javascript.
- 4) Hold the pre1st/2nd/3rd test. For the pre test, the goal is to get the draft workflow of interfaces fast. For the 1st & 2nd test, the goal is to assess the usability of interfaces and to explore what designers want and how they want to interact with the platform more. For the 3rd test, the goal is to assess if the output meet the needs for general cases by testing different use cases.
- 5) Analyze results got from tests. Then iterate prototypes based on insights from analyzing.

Goal: Develop an interface for model-informed prototyping using HAI communicative acts as design materials. The interface will be used as a prototyping tool for Human-AI interactions. The tool will be used to collect a set of use cases, models, users, as well as designed interactions for different scenarios. Through workshop sessions, the goal is to develop a tool which supports designers to follow a set of guidelines while prototyping HAI interactions.

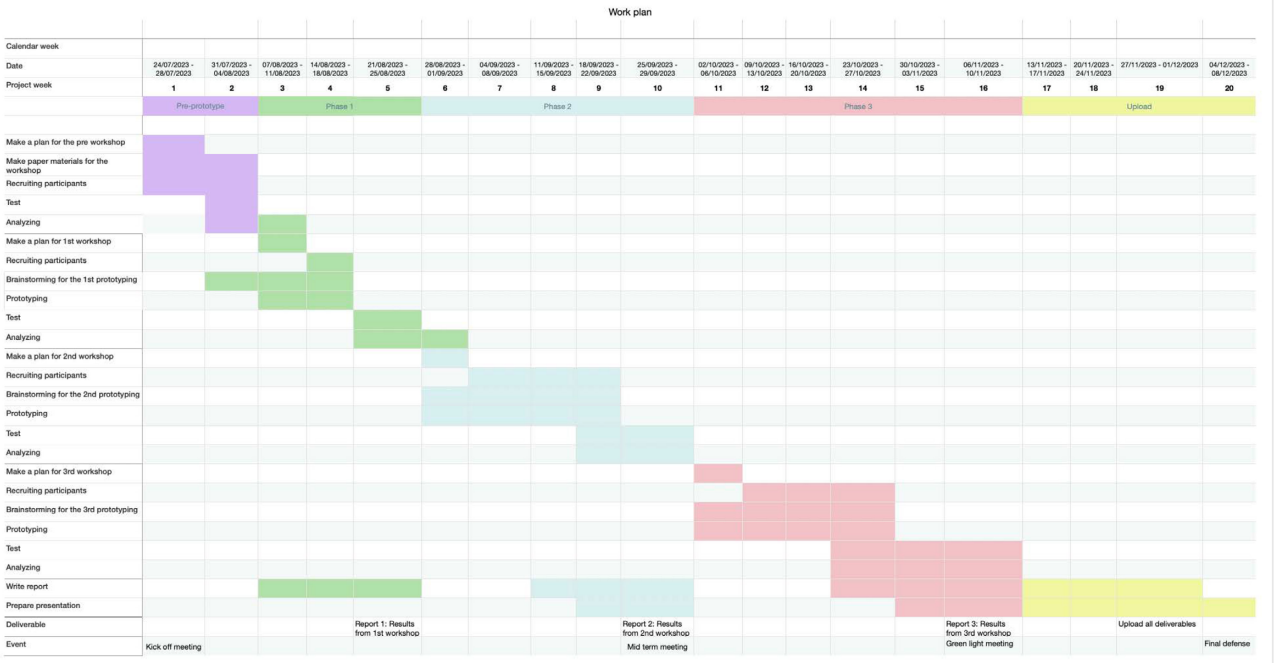
Initials & Name J Zhou Student number 5568935

Title of Project Develop an interface for model-informed prototyping of HAI interactions

PLANNING AND APPROACH **

Include a Gantt Chart (replace the example below - more examples can be found in Manual 2) that shows the different phases of your project, deliverables you have in mind, meetings, and how you plan to spend your time. Please note that all activities should fit within the given net time of 30 EC = 20 full time weeks or 100 working days, and your planning should include a kick-off meeting, mid-term meeting, green light meeting and graduation ceremony. Illustrate your Gantt Chart by, for instance, explaining your approach, and please indicate periods of part-time activities and/or periods of not spending time on your graduation project, if any, for instance because of holidays or parallel activities.

start date 24 - 7 - 2023 8 - 12 - 2023 end date



My project planning phase is based on an Iterative prototyping method (Figure 2). There are 3 phases in total. For each phase, I describe the corresponding activities as follows:

Pre/Phase 1 -> 1) Make a plan on what the test (workshop) is, including participants features, tasks, questionnaires and other material we need; What's more, we also need to make one standard for assessing if the prototype meets the design considerations above. 2) Invite participants and make appointments. 3) Using Figma and/or paper material for the prototyping building. 4) Hold the test and analyzing insights from results; During the test, I would ask participants Think aloud" and use the video to record the whole process while observing what they do. After that, I plan to interview them based on results they "design" with the low-fi prototypes. 5) Evaluate the prototype into a better digital prototype and prepare for the next test. -> After these phases, I would get insights on the workflow, the frame, the function and most details on how designers could interact with the platform to design for the end users.

Phase 2/3 -> 1) Make a plan on what the test (workshop) is, including participants features, tasks, questionnaires and other material we need. 2) Invite participants and make appointments. 3) Using html/CSS/Javascript as the programming tools to create the final prototype. P5.js would be another tool to add visual elements if necessary. 4) Hold the test and analyzing results. -> After this phase, I would output the final prototype which should be useful in more use cases.

Delivery: UI/UX design; create, test and evaluate the final prototype, Report 1& 2 & 3 -> evaluate the final design output.

Initials & Name J Zhou Student number 5568935

Title of Project Develop an interface for model-informed prototyping of HAI interactions

MOTIVATION AND PERSONAL AMBITIONS

Explain why you set up this project, what competences you want to prove and learn. For example: acquired competences from your MSc programme, the elective semester, extra-curricular activities (etc.) and point out the competences you have yet developed. Optionally, describe which personal learning ambitions you explicitly want to address in this project, on top of the learning objectives of the Graduation Project, such as: in depth knowledge a on specific subject, broadening your competences or experimenting with a specific tool and/or methodology, ... Stick to no more than five ambitions.

Programming Skills: Learn in-depth programming knowledge about artificial intelligence to create an interactive and explainable AI product. Programming languages such as Python are widely used in modern society and can be used to create a wide range of digital art and interactive experiences through in-depth knowledge of them, providing a reliable means of implementing design outcomes.

User Experience Design ability: One of the most critical aspects of designing interpretable AI products is creating a transparent and efficient user learning experience. Learning UX design principles and best practices can help me implement design ideas that are intuitive and easy to use for customers, which can improve their overall satisfaction and help AI education grow.

Human-AI Interaction Design: Based on my current career plan, I expect to gain more knowledge and practical experience about HAX through this final project, and look for design possibilities in artificial intelligence products.

Reference:

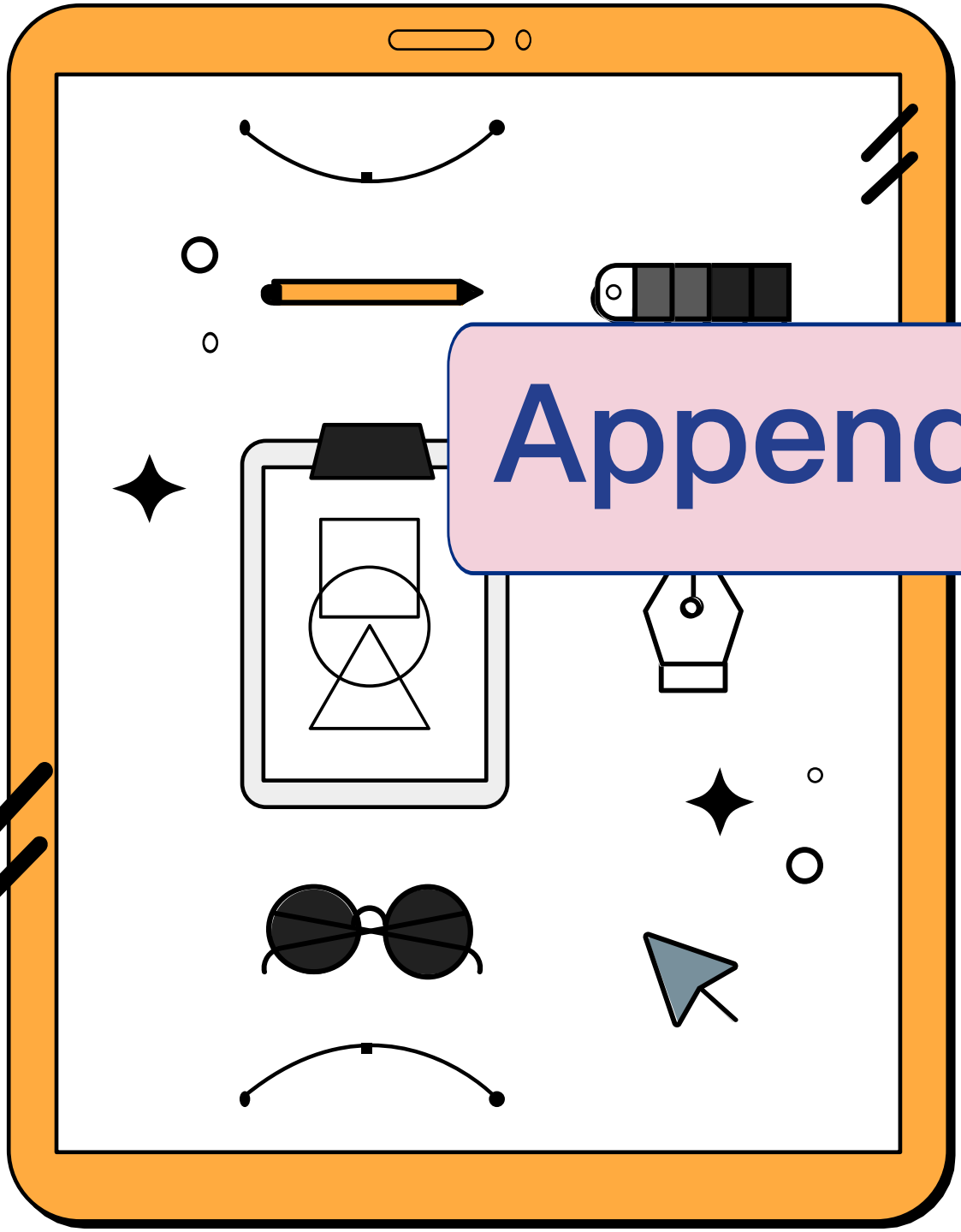
[1] Russell, S. (2022). If We Succeed. Daedalus, 151(2), 43–57. https://doi.org/10.1162/daed_a_01899.
[2] Subramonyam, H., Seifert, C., & Adar, E. (2021, June). Towards a process model for co-creating AI experiences. In Designing Interactive Systems Conference 2021 (pp. 1529-1543).
[3] Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–13.
[4] Subramonyam, H., Seifert, C., & Adar, E. (2021, April). Protoai: Model-informed prototyping for ai-powered interfaces. In 26th International Conference on Intelligent User Interfaces (pp. 48-58).
[5] Google. 2019. People + AI Guidebook. https://pair.withgoogle.com/
[6] Apple Inc. 2019. Designing the UI and User Experience of a Machine Learning App. https://developer.apple.com/design/human-interface-guidelines/machinelearning/overview/introduction/
[7] Goldman, N., & Narayanaswamy, K. (1992, June). Software evolution through iterative prototyping. In Proceedings of the 14th international conference on Software engineering (pp. 158-172).

FINAL COMMENTS

In case your project brief needs final comments, please add any information you think is relevant.

Initials & Name J Zhou Student number 5568935

Title of Project Develop an interface for model-informed prototyping of HAI interactions



Appendix B

NEW SAMPLE

a new sample

Sample (Input to model):
an instance from data, e.g., image, sound, form, video

Users can generate new samples by uploading a new file, capturing an image, etc.

Models can create a new sample, e.g., random feature values

Examples — capture a new face, create a new document, create a new feature vector

SAMPLE FOR LABEL

sample for a given label

Sample (model input):
an instance from data, e.g., image, sound, form, video

Label (model output):
a class/category which characterizes the sample, e.g., face, utterance, pass/fail, etc.

Users can select an existing sample or create a new one, which belongs to the given label

Models can select a sample - its prediction using this sample should be the provided label

Examples — categorize a document, detect an activity, classify an image, predict success

EDITED SAMPLE

a sample modified to a new sample

Sample (model input):
an instance from data, e.g., image, sound, form, video

Users can modify an existing sample, through the interface, e.g., change features, edit file, etc.

Models can modify an existing sample by altering the features of the feature vector or the raw data

Examples — modified document, edited image, changed feature value

SAMPLE FROM LIST

sample selected from a list of samples

Sample (model input):
an instance from data, e.g., image, sound, form, video

List of samples (model input):
a list of samples is a list of model inputs, e.g., list of files, forms, feature vectors, etc.

Users can select a sample through a list based on the interface, e.g., click

Models can select a sample based on the interface of the interaction, e.g., highlight selected sample on a visualization

Examples — select a file, choose a feature vector, click on a data point (visualization)

EDITED SAMPLE FROM PREDICTION

modified sample from a sample- label pair

Sample (model input):
an instance from data, e.g., image, sound, form, video

Label (model output):
a class/category which characterizes the sample, e.g., face, utterance, pass/fail, etc.

Users can modify a sample through the interface of the interaction, e.g., change features, edit file, etc., in order to match the label

Models can modify a sample through the interface, by altering the features of a feature vector or editing an image, in order to match the label

Examples — modified document, edited image, changed feature value - to match the label

LABEL FOR SAMPLE

label for a given sample

Sample (model input):
an instance from data, e.g., image, sound, form, video

Label (model output):
a class/category which characterizes the sample, e.g., face, utterance, pass/fail, etc.

Users can annotate a sample by assigning a label to it (human annotation)

Models can predict the label of a sample through the model predict function

Example — categorize a document, detect an activity, classify an image, predict success

SAMPLE AND LABEL

a sample- label pair

Sample (model input):
an instance from data, e.g., image, sound, form, video

Label (model output):
a class/category which characterizes the sample, e.g., face, utterance, pass/fail, etc.

Users can provide an example - a labeled sample through the interface; upload or select a sample with its label

Model can communicate a labeled sample through the interface, e.g., visualize a sample and its predicted/provided label

Examples — annotated image, positive/negative example, new data point with label

EDITED LABEL FROM PREDICTION

edited label from a sample-label pair

Sample (model input):
an instance from data, e.g., image, sound, form, video

Label (model output):
a class/category which characterizes the sample, e.g., face, utterance, pass/fail, etc.

Users can edit the label of the sample through the interface, e.g., click new label

Model can modify the label by predicting a different class for the given sample

Examples — change the prediction for a detected face, alternate class for document

PREDICTION PROBABILITY

the probability for a sample-label pair

Sample (model input):
an instance from data, e.g., image, sound, form, video

Label (model output):
a class/category which characterizes the sample, e.g., face, utterance, pass/fail, etc.

Probability (model output):
the probability (0-100%) that the sample belongs to the given label.

Users may estimate the probability of a label being true

Models can estimate the probability of a given sample being classified into the specific label

Examples — the probability of a spam email, the probability of a document being accepted

GLOBAL EXPLANATION

explanations for the model functionality

Explanation (global XAI):
explanations of how the model makes predictions, e.g., which features are most important, which rules the model follows to make predictions, etc.

Users can provide high-level explanations for the model decisions, e.g., rule-based explanations, ranking features based on their importance, etc.

Models can provide explanations about their functionality by visualizing graphs (feature importance values, in natural language (rule-based explanations), etc.

Examples — rules, graphs, ranking features based on their importance for the model's predictions

LOCAL EXPLANATION

explanation for a given sample- label pair

Sample (model input):
an instance from data, e.g., image, sound, form, video

Label (model output):
a class/category which characterizes the sample, e.g., face, utterance, pass/fail, etc.

Local explanation
a justification for the given sample-label pair - why this prediction was made

Users can justify their decision through if-then rules (rule-based explanations)

Models can provide explanations about a given prediction through feature importance values, rule-based explanations, etc.

Examples — validate the recognition of an object, agree with a categorization of a document

FEEDBACK FOR PREDICTION

validation feedback for a sample- label pair

Sample (model input):
an instance from data, e.g., image, sound, form, video

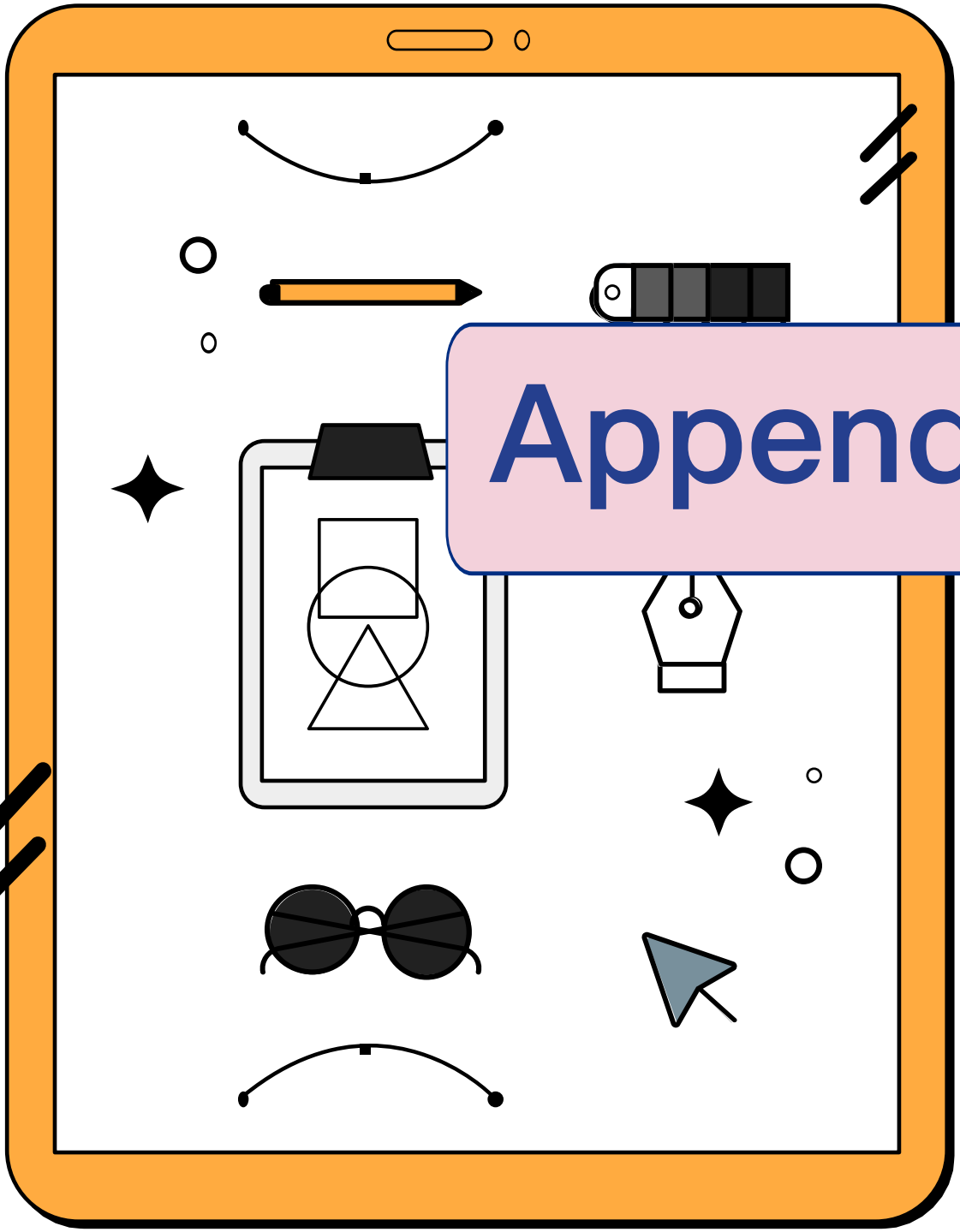
Label (model output):
a class/category which characterizes the sample, e.g., face, utterance, pass/fail, etc.

Validation feedback
a "Yes" or "No" signal which describes the (dis)agreement with a given prediction (label for sample)

Users can provide validation through clicks, buttons, gestures, etc. .

Model can provide a validation based on their own prediction for the sample

Examples — validate the recognition of an object, agree with a categorization of a document



Appendix C

U2: Applicant

Description: A (prospective) applicant who uses the model to assess or apply with their CV

- **Use Case:** CV recruitment
- **Role:** Applicant
- **Expertise:** None

Example of interactions

- uses model to apply for a position
- uses model to understand how it makes decisions
- uses model to assess and improve their CV

U1: Manager

Description: A manager who uses the model to make decisions about the hiring process

- **Use Case:** CV recruitment
- **Role:** Manager
- **Expertise:** Domain (hiring) -- No AI

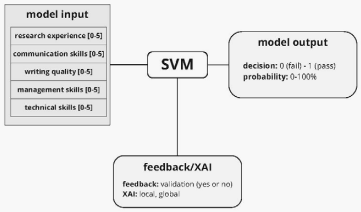
Examples of interaction

- uses model to make decisions about applicants
- uses model to understand how it makes decisions
- adjusts model based on new data or decisions

Model M: CV-scoring

A supervised ML model which predicts if a CV will be shortlisted or not, based on applicant's skills for a specific job application

- **Model type:** Support Vector Machine (SVM)
- **Model task:** Classification



- **Input:** CV as a feature vector (skills)
- **Output-Class:** Shortlisted or Not (0-1)
- **Output-Metric:** Probability of CV being shortlisted (or not)
- **Feedback/XAI:**
  - **local explanation:** explain prediction
  - **global explanation:** explain model
  - **validation:** validate prediction (agree or not)

Description:

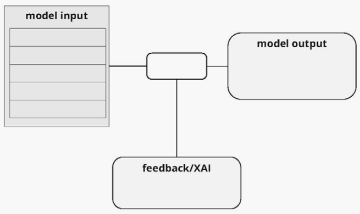
- **Use Case:**
- **Role:**
- **Expertise:**

Example of interactions

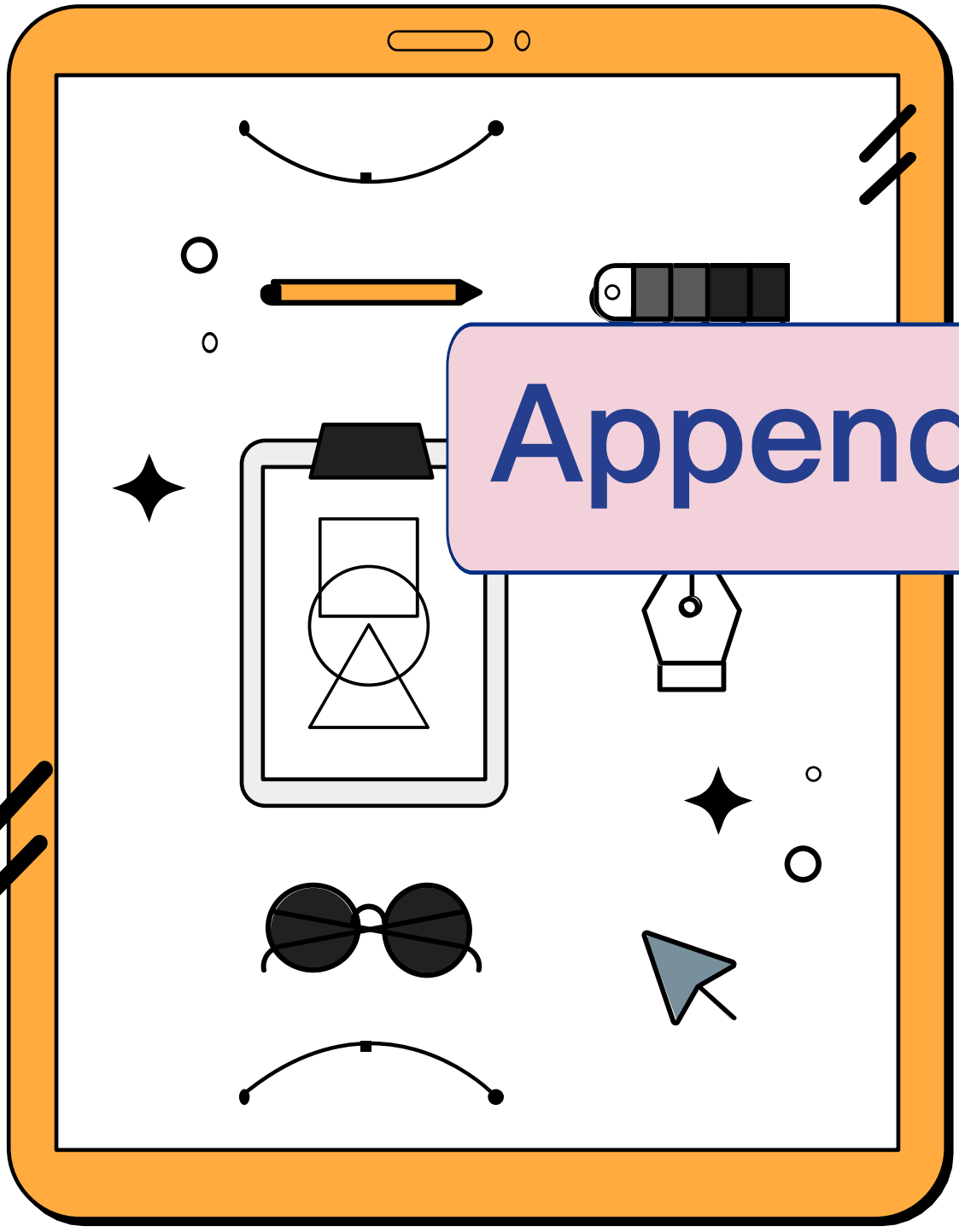
-
-
-

Model __ : _____

- **Model type:**
- **Model task:**



- **Input:**
- **Output-Class:**
- **Output-Metric:**
- **Feedback:**
  - **(local) explanation:**
  - **(global) explanation:**
  - **validation:**



Appendix D

PROCEDURE

What are the main steps in the test procedure?

0 - 2 Min Welcome the participant	Introduce the test and ourselves. Ask user consent!
0 - 2 Min Check the prototype	Make sure that everything is ready. Check if the recording works for tracking the test activity.
2 - 5 Min Pre test interview	General background questions to get a better picture of the participant.
15 - 20 min Conducting tasks	During this part, show them the context. Then participants do Activity 0 - 1.
15 - 20 Min Order messages	Ask participants to do Activity 2.
15 - 20 Min Post test interview	Interview about why participants choose the order or sequence.
2 - 5 Min Test sum up	Organize notes and come up with insights and problems.

LOCATION & DATES

Where and when will the test take place? When and how will the results be shared?

The test is conducted at IDE in Delft. Each test will be conducted with one participant at a time. One test would be held in Wednesday, the other one would be held in Sunday.

EQUIPMENT

What equipment is required? How will you record the data?

- Tools to help record the test include, paper & pen, camera, computer and charger.
- Paper model and explanation.
- Camera for making pictures.
- Consent forms.
- Chairs and tables.

TEST TASKS

What are the test tasks?

Context: Understand concepts
Activity 0: The activity is to test if the participant understands what message means and how it is made of.

Context: Apply for a job
Activity 1: The activity is ask participants to think about how the second times the process would be like.
Activity 2: The activity is ask participants to order the messages into Human-AI interactions based on the path which makes sense for designers.

PARTICIPANTS

How many participants will be recruited? What are their key characteristics?

2 participants and they would be students majoring in design.

INTRODUCTION

"Hi, welcome to this experience, I am Jiayi Zhou and thank you for participating in the test. The test and interview afterwards will be recorded, the video is only used for me and the project team analysis and will be deleted after the project is finished. By signing the consent form you will agree with this. The test consists of three parts: First I will ask you some questions to know your basic information. Then we will provide you with materials and introduction on concepts used in the test and ask you to use paper cards to complete activities. If possible, it is better for you to think out loud while performing. And if there is any question, you could ask me for explanation or help. I would help you if the question doesn't affect the test. At the end of the test I will have a short interview with you. You could give me any feedback, positive or negative both are very important. Do you have any questions?"

PRE-TEST QUESTIONS

What basic information the test needs?

1. What's your name?
2. Are you a design student?
3. Do you have experienced in design or get an experience in design education?
4. Do you have experience on AI and design area?

AFTER TEST QUESTIONS

What questions would be asked after the test?

1. Why do you write the Scenario 2 like this?
2. Why do you choose this sequence for the Scenario 1?
3. Why do you choose this sequence for the Scenario 2?
4. Why do you add/remove this message from the sequence?
5. If there is any more interesting questions showing in the test, take a note and ask here.

Prototyping Human-AI Interactions - Pre

Objective: test if participants understand what messages mean, what the context is and what the interactions would be.

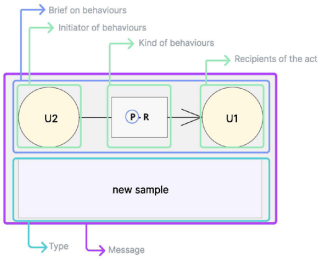
Materials

- user & model cards
- types cards
- message templates (to be filled-in)

Task Description

Below, there are some textual descriptions. Based on each description you should fill the brief on behaviours and then assemble the right message card.

e.g. applicant submits a CV to the manager



- applicant submits their CV to the manager
- manager asks the model why it classified a specific CV as unsuccessful
- manager prompts the user to provide an updated CV
- manager validates a model's decision for a provided CV
- applicant wants the manager to show a successful CV
- manager asks the model to change its decision for a provided CV
- applicant asks the model to change the decision for their CV
- model prompts the manager to provide a scored CV
- manager asks the model to make decision for a provided CV
- manager informs the applicant about the decision for their CV

Prototyping Human-AI Interactions - Activity 1

Objective: explore how participants think of the possible interactions in a new context.

Materials

- user & model cards
- types cards
- scenario 1 card
- scenario 2 card
- pen
- sticky notes

Task Description

Referring to Scenario 1, fill the blank by words or pictures in Scenario 2 to organize the story between actors (the U1, the U2, the Model).

Scenario 1

The U2 submits an CV to the U1 for the first time.

The U2 submits their CV #1to the U1. The U1 uses the Model to get a prediction for the CV #1. The Model gives the U1 the prediction #1, but the U1 gives a negative feedback to the prediction#1 and asks the Model to give explanation on why the Model made the specific decision and how the Model works. Based on the negative feedback, the Model gives the U1 the prediction #2. The U1 gives a positive feedback to the prediction #2. Then the U1 informs the U2 the prediction #2.

Scenario 2

The U2 submits an CV to the U1 for the second time.

After the U1 informs the U2 the prediction #2, the U2 edits their CV #2 and submits it to the U1 again. The U1 uses the Model to get a prediction for the CV #2 again.

The U1 gives a positive feedback to the prediction #2. Then the U1 informs the U2 the prediction #2.

Prototyping Human-AI Interactions - Activity 2

Objective: explore the structure of the path how messages pass on in different context.

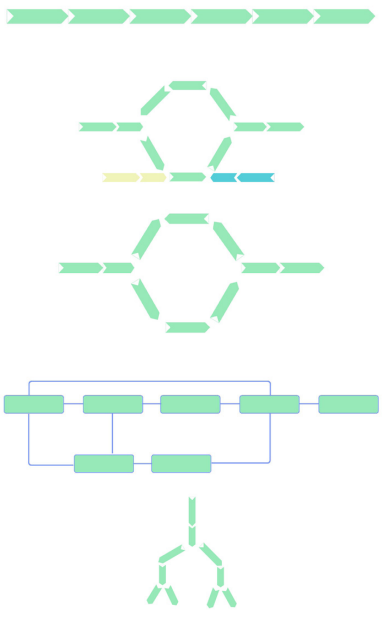
Materials

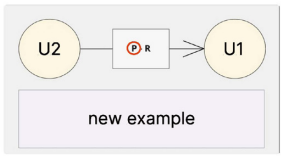
- user & model cards
- types cards
- scenario 1 card
- scenario 2 card
- pen
- messages (fillen-in)
- message template (to be filled-in)

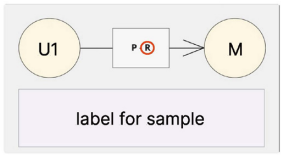
Task Description

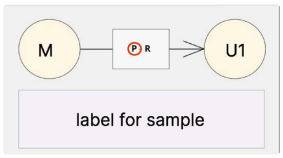
Based on the textual descriptions of scenarios, you have to use the pre-filled messages to formulate an interaction as a sequence of messages in order to describe te given scenario. **Make sure the sequence makes sense for you. And you could freely add any message.**

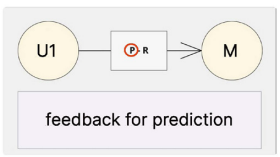
Here are some messaging structures you can refer to. You can use them to help you construct sensible messaging paths. However, you could use other ways to make it!

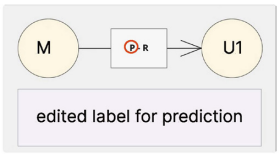


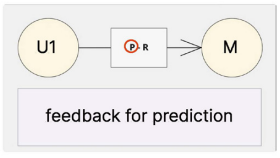


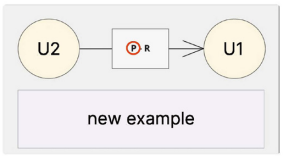


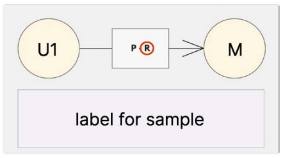


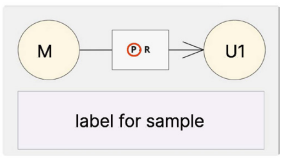


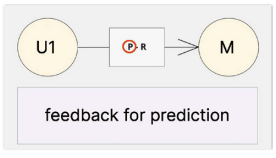


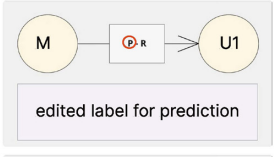


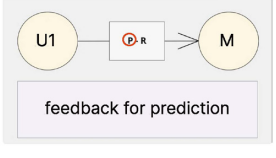


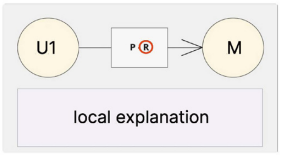


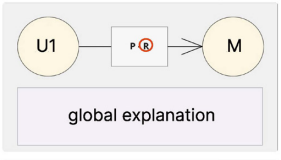


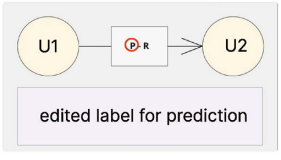


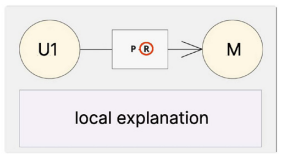


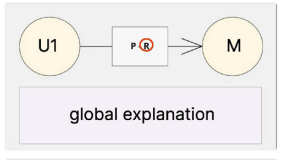


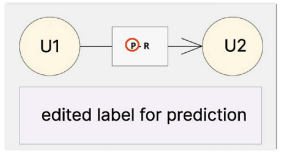


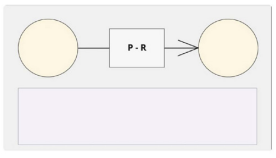


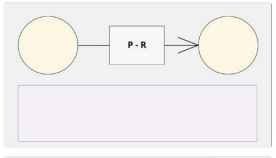


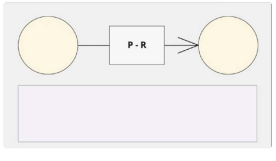












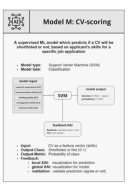
front

Instructions - Description of Materials

- Human-AI interactions are interactions between **human users** and **AI models**
- Users and models can exchange model-based **types of information**

- **User Cards:** describe (a type of) user based on their role in the interaction and their AI/domain expertise

- **Model Cards:** describe the model used in the interaction, including the model type and the model-specific types (input, output, feedback)



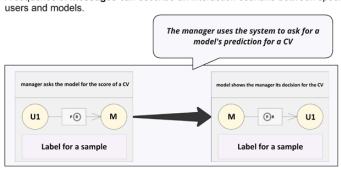
back

Instructions - Description of Materials

- **Types of information:** There are 13 types of information that can be exchanged between users and models

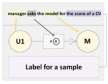
- A **message** describes the communication between a sender and a receiver in the context of a use case (*User and Model cards*). The sender can *provide* a type to the receiver or *request* a *type from the receiver*

- A sequence of **messages** can describe an interaction scenario between specific users and models.



Definitions and terms

- **Users** and **models** interact with each other using **messages**



- **Messages** are used to **provide** or **request** information
  - The sender (user or model) can provide information to the receiver (user or model)
  - The sender (user or model) can request information from the receiver (user or model)
- **Messages** use **communicative acts** to describe the **type of information:**
  - Communicative acts are defined based on what type of information a model can provide based on its specifications, i.e., model input/output, model metrics, explanations, and others.

  - We provide a representation of such communicative acts using the cards

List of types

a data type (given another data type)

- ***new sample***: a new sample
- ***sample from list***: sample selected from a list of samples
- ***sample for label***: sample for a given label
- ***edited sample***: a sample modified to a new sample
- ***edited sample for prediction***: modified sample from a sample-label pair
- ***sample and label:*** a labeled sample
- ***label for sample***: label for a given sample
- ***prediction probability***: the probability for a given sample-label pair
- ***edited label for prediction:*** alternate label for a given sample-label pair
- ***local explanation***: explanation for a given sample-label pair
- ***global explanation***: explanation for model predictions
- ***feedback for prediction***: validation feedback for a given sample-label pair

new sample

a new sample

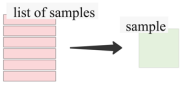
- **sample: model input**
  an instance from input data, e.g., image, sound, form, video, etc.

- the new sample is in the model input format but it is probably new to the model
- users generate new samples by uploading a new file, capturing an image, etc.
- models can create a new sample, e.g., random feature values
- **Examples:**
  capture a new face, create a new document, create a new feature vector

sample from list

sample selected from a list of samples

- **sample: model input**
  an instance from data, e.g., image, sound, form, video, etc.
- **list of samples: model input**
  a list/collection of samples

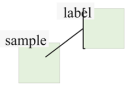


- user can select a sample through the interface of the interaction, e.g., click
- models can select a sample based on the interface of the interaction, e.g., highlight selected sample on a visualization
- **Examples:**
  select a file, choose a feature vector, click on a data point (visualization)

sample and label

sample and label

- **sample: model input**
  an instance from data, e.g., image, sound, form, video, etc.
- **label: model output**
  a class/label which characterizes the sample, e.g., face, utterance, pass/fail, etc.

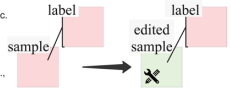


- users can provide an example - a labeled sample through the interface; upload or select a sample with its label
- models can communicate a labeled sample through the interface, e.g., visualize a sample and its predicted/provided label
- **Examples:**
  annotated image, positive/negative example, new data point with label

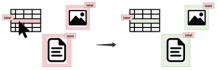
edited sample from prediction

change the sample from a prediction

- **sample: model input**
  an instance from data, e.g., image, sound, form, video, etc.
- **label: model output**
  a class/label which characterizes the sample, e.g., face, utterance, pass/fail, etc.



- users can modify a sample through the interface of the interaction, e.g., change features, edit file, etc., in order to match the label
- models can modify a sample through the interface, by altering the features of a feature vector or editing an image, in order to match the label
- **Examples:**
  modified document, edited image, changed feature value - to match the label

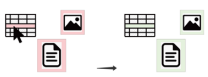


edited sample

change sample to a new sample

- **sample: model input**
  an instance from data, e.g., image, sound, form, video, etc.

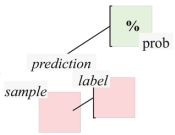
- users can modify a sample through the interface, e.g., modify file, change feature values
- models can select a sample based on the interface of the interaction, e.g., highlight selected sample on a visualization
- **Examples**:
  modified document, edited image, changed feature value



prediction probability

the probability for a given prediction

- **sample: model input**
  an instance from data, e.g., image, sound, form, video, etc.
- **label: model output**
  a class/label which characterizes the sample, e.g., face, utterance, pass/fail, etc.
- **probability: model output**
  the probability (0-100%) that the sample belongs to the given class.

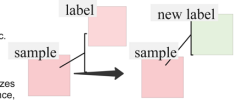


- users may estimate the probability (level) of a prediction being true through the interface
- models can estimate the probability of a given sample being classified into the specific class
- **Examples:**
  the probability of a spam email, the probability of a document being accepted

edited label for prediction

edited label for a given prediction

- **sample: model input**
  an instance from data, e.g., image, sound, form, video, etc.
- **label: model output**
  a class/label which characterizes the sample, e.g., face, utterance, pass/fail, etc.



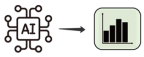
- users can edit the label of the sample through the interface, e.g., click new label
- models can modify the label by predicting a different class for the given sample
- **Examples:**
  change the prediction for a detected face, alternate class for document

explanation for the AI model

global explanations for model functionality

- **explanation: global XAI**
  the graph visualizes how the features affect the model's decisions

- the model can compute and visualize the feature importance values based on which it makes predictions
- users can provide an estimation of a ranking based on the importance of the features to the model's decisions
- **Examples:**
  rules, graphs, ranking based on feature importance for the model's predictions

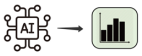


explanation for the AI model

global explanations for model functionality

- **explanation: global XAI**
  the graph visualizes how the features affect the model's decisions

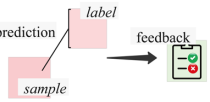
- the model can compute and visualize the feature importance values based on which it makes predictions
- users can provide an estimation of a ranking based on the importance of the features to the model's decisions
- **Examples:**
  rules, graphs, ranking based on feature importance for the model's predictions



feedback for prediction

validation feedback for a given prediction

- **sample: model input**
  an instance from data, e.g., image, sound, form, video, etc.
- **label: model output**
  a class/label which characterizes the sample, e.g., face, utterance, pass/fail, etc.



- **validation/feedback**
  a "Yes" or "No" signal which describes the (dis)agreement with a given prediction (label for sample)
- users can provide validation through clicks, buttons, gestures, etc. .
- models can provide a validation based on their own prediction for the sample
- **Examples:**
  validate the recognition of an object, agree with a categorization of a document

label for sample

label for a given sample

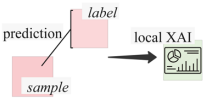
- **sample: model input**
  an instance from data, e.g., image, sound, form, video, etc.
- **label: model output**
  a class/category which characterizes the sample, e.g., face, utterance, pass/fail, etc.

- users can annotate a sample by assigning a label to it - human annotation
- models can predict the label of a sample through the model predict function
- **Examples:**
  categorize a document, detect an activity, classify an image, predict success

local explanation

explanation for a given prediction

- **sample: model input**
  an instance from data, e.g., image, sound, form, video, etc.
- **label: model output**
  a class/label which characterizes the sample, e.g., face, utterance, pass/fail, etc.

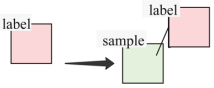


- **local explanation**
  the importance of the features based on their role in the specific prediction
- the model can compute and visualize the feature importance values for a prediction
- users cannot estimate such importance values - users can provide justification for a prediction, e.g., if-then rules
- **Examples:**
  show the features that were more responsible for the prediction

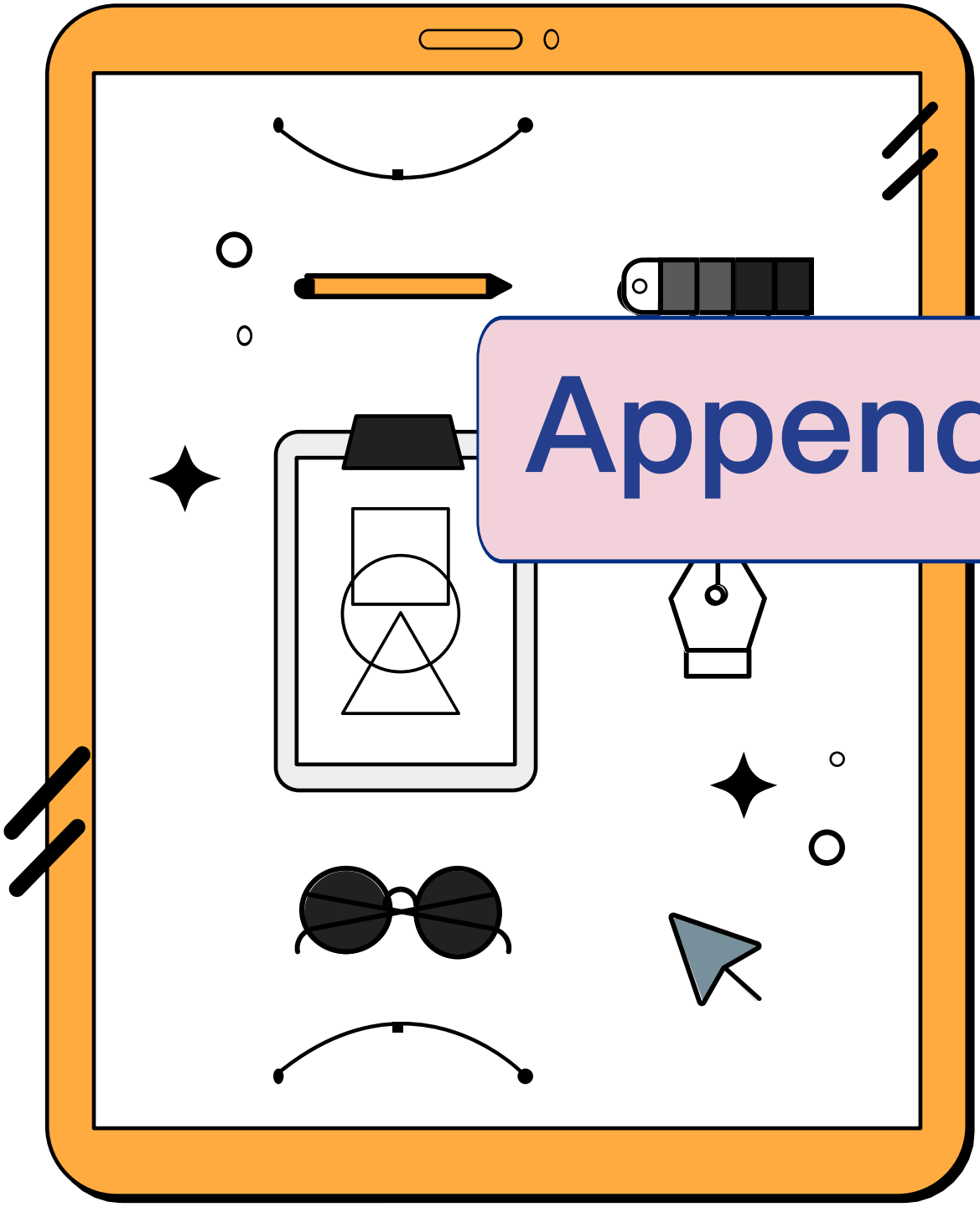
sample for a label

sample for a given label

- **sample: model input**
  an instance from data, e.g., image, sound, form, video, etc.
- **label: model output**
  it characterizes the sample, e.g., face ID, utterance, pass/fail, etc.



- users can select an existing sample or create a new one, which belongs to the given class
- models can select a sample - its prediction using this sample should be the provided label
- **Examples:**
  categorize a document, detect an activity, classify an image, predict success



Appendix E

Transcription for Participant 1

Pre-test interview

P1
Role: A student majoring in Design for Interaction(DFI) currently in TUD
Design/design education experience: More than 10 years
Experienced in design & AI: ITD

Activity - Pre

■ Participant 1 ■ Researcher

1. "I don't understand what's the meaning of 'P' or 'R'?" " 'P' means 'Provide', 'R' means 'Request. " "Make sense."
2. "It says 'why' and 'specific CV', so it should be explanation in the context and to the CV."
3. "So from the third sentence, it means the manager wants the applicant would pass the model assessment, right? In that case, I think there is a note between the computer and the manager's permissions." "Forget the permissions of the computer and the manager, it is just an ideal description."
4. "This sentence emphasizes 'validate' and 'decision', so it is the feedback."
5. "Why do you think the type of information is 'sample and label'?" "It is called 'a successful', so I think it must be labeled by the model. " "What's the difference between the 'Sample for label' or 'Sample and label' for you?" "To be honest, for the sentence, I think they are the same. However, if I have to see the difference, 'Sample for label' means we use one label to search for the responding example while the other one means we need to find a sample with the specific label. In other words, we don't use the label as a cue to find the sample." "After reading the details about the both types of information, do you still think the sentence is 'Sample and label'? " "Yes, I think there is no difference."
6. "Here, the same as the 3rd one. Manager would be the one make the final decision, right?"
7. "I think it is the same as the last one instead of the actor 1."
8. "What's the meaning of scored? Is it the same as the successful? " "It means the percentage of results you think this prediction is right. " "I find it a bit strange if "scored" means the final result. For example, there are 100 points in total, and this CV has 60 points. What's more, who gives the CV the score? The computer or the manager? "
9. "The left are easy for me. "

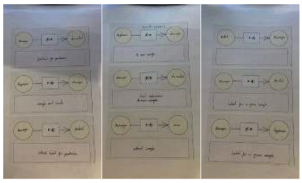


Figure 0. Participant 1's pre

Permissions between the U1 and the model

Who gives the label?

Permissions between the U1 and the model

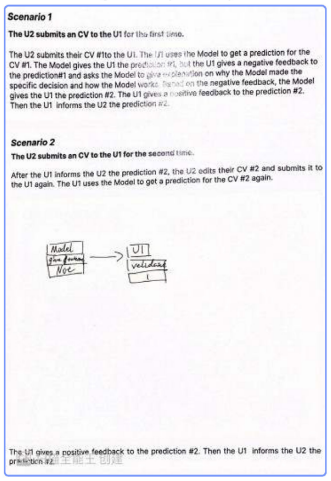
Who gives the label "scored"? What's the 'scored' mean?

Activity - 1

■ Participant 1 ■ Researcher

1. "Firstly, I have some questions on the Scenario 1. What's the meaning of positive or negative feedback? Are they the feedback from the manager to the model or the manager to the applicant?" After reading the scenario again, "Ok, now I understand what they mean."
2. "I think the story for the Scenario 2 is flexible because it could be a long story or a very short story. " "It depends on which is the most logical story for you." "So it would be a very short story like this (Figure 1)."
3. "Why do you think the story makes sense? Or could you explain the story?" "After the manager uploads the new edited CV, for the computer, it has one stored feedback. I don't think it would change the feedback, also known as the label, just based on some description changing. At the same time, I don't think the model would ask the manager for more feedback on the same person's CV again because in the first time it has a lot of data on feedback and it should learn how to make the right decision. "
4. "I also have to say, we need to think about the rights of the manager and the model. How much impact can a manager's feedback have on a decision or label output by a model? 100%? 90%? Or less?"
5. "The role of machines is also very important to think about. Is the machine a tool to assist or is it solely responsible for the CV screening process? In general, if the manager doesn't know the applicant at all, they may just take the model's label or suggestion. If the manager really appreciates the applicant, but for some other reason, the model doesn't give the applicant's CV a feedback that satisfies the manager, then the manager may keep asking the model to make changes."
6. "Why do you think the edited CV cannot be successful in the second time again? " "I don't know. It is the model's work. But I am sure it has its own assessment. But it doesn't matter."

Figure 1. Participant 1's story



Feedback based on the last one

Model evaluates from the last feedbacks

Permissions between the U1 and the model

Edit the model's data

Activity - 2

Figure 2. Sequence fo the Scenario 1 of Participant 1

Figure 3. Sequence fo the Scenario 2 of Participant 1

Specific needs of users

Different roles means different sequence

Permissions between the U1 and the model

Who gives the label?

Activity - 2

■Participant 1 ■ Researcher

Figure 2

1. "The sequence is based on how much granularity you want. For example, in the 'model requests u1 to give feedback on the prediction' and 'u1 gives the feedback to the prediction', here might be two choices. " "You could write down the most logical one." "It is hard to say because it depends on the specific requirements of the user's position. Again, for me, I could image when the applicant provides the system or maybe the model with the CV and then the manager requests the model to give a feedback on the CV. Of course, sometimes the manager needs to input the CV into the model by hand. What' more, again, how is decision-making power allocated? The distribution of decision making power and the way things are done affects the sequence or order of the information. i think a more specific description is needed to know what sequence should be taken."
2. "For example, if I know this applicant and I already have a subjective judgment about this person, how do my subjective judgment and the model's objective judgment affect the final decision, respectively? Can I change the overall result subjectively? It's also possible that I don't know each candidate and just want the machine to give me the final result."
3. "If it is accessible for the manager to input something which might affect the model's prediction, how do you do that? What would you choose?" "I think the manager would set the labels or standards in the beginning or in the middle or somewhere else." See Figure 4. "I would call them new labels. And these labels are made by the manager instead of the model. And it might be given for a specific opinion or just in general. The sequence is flexible based on the specific case."

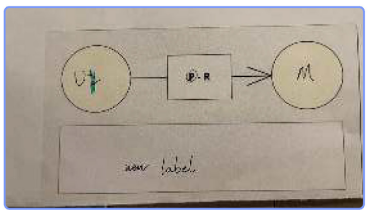


Figure 4. Set new labels for the assessment

Interfaces affect the sequence.

Flexible process

Explanations will be asked for results that do not meet expectations, and vice versa.

Permissions between the U1 and the model

Activity - 2

■Participant 1 ■ Researcher

4. "While we have to think about the sequence on the local explanation, for me, it depends on the interface. If it is a button, I would think the message would be like Figure 5. If it is a chat box, the manager could input the detailed question, it would be like Figure 5. Now the new sample means the details about the question."
5. "Why do you think the sequence is like this?" "Because the story is every linear. And it is the fast way to solve problems. The applicant provides the manager with the CV and then the manager requests the model to give the label. However, if it is an automatic process, I think it should be the model provides the manager with the label automatically instead of one more step 'the manager requests the model to label for sample.' Then the manager would decide if the label is right. After that, the manager could ask the model why it provides the result. " "Why do you think the manager would ask about it after the feedback for the label?" "Yes it could also be like that. But in my sequence, it is not solid workflows which every step is fixed and we have to do the first one and then the next. We could change the order by ourselves. If the manager says yes to the label, they won't ask any question. However, if they say no, there might be some chances they want to know the local or global explanation. And the reading material also leads me to do that."
6. "If there are more details, how do the sequence change in the 'edit label for the prediction' part? " Figure 6. "It depends on the how the manager and the model could affect the final result. While I am ordering the message, the model works like an assistant. That means it is the manager who makes the final decision. Of course, maybe there would be another buddy having more powder who can change everything and valid the manager's feedback. However, if the model could work independently, it could change the label based on the manager's feedback."

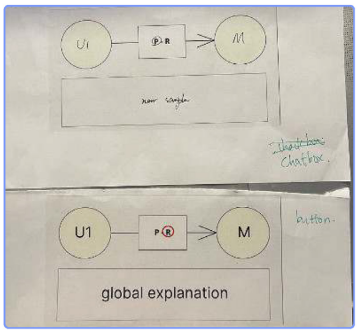


Figure 5. Different interfaces on message

Figure 6. More details on the message

Permissions between the U1 and the model

The goal of effective workflow -> quick & short

Roles of the actors & model affect the workflow

Motivation of asking the explanat ion

Activity - 2

■Participant 1 ■ Researcher

7. "If there is one pop-up given by the model to ask the manager's feedback, what do you think of it?" "For the user experience, it is hard to say if it would develop it or not. For me, I would be curious why the model asks me this question or gives me the pop-up? If the model works independently, why does it need my suggestion? Is it to develop its machine learning result? I don't think it is related to the workflow. What's more, in my story, the model is not independent. The manager uses the model to assess the CV, now the model asks the manager questions and it is a little strange."
8. "I want a workflow that helps me make quick decisions. There would be two situations on the local explanation or global explanation. The first is the manager scans the CV and has one opinion, then uses the model to see if there is other suggestion. Then the manager could get one more opinion from the other side. And if the model's opinion is the same as the manager's one, the manager probably won't ask for the both explanations. It requires a quick and short workflow. However, if the model's opinion is different, the manager might want to ask some questions. To be honest, this part is not connective strongly in the whole process to achieve the goal. It seems more like the manager works for the model to help it develop instead of getting the decision for the CV. And it feels the manager is discussing something with the model. However, from the process, I don't think the manager is communicating with the model. It is only the requesting-providing relationship. If the manager doesn't have a look at the CV, it is possible the manager has no feedback or needs to ask for any explanation. And in this situation, the applicant could upload their CV to the model directly. Then the

What causes the manager think they need the help of
9. "What does the model bring us? " "If you need to ev workflow, what would you do?" "Image the whole si the first one. Then the applicant provides the mang the manager has a look at it and then requests the label. This label means opinion from the other side manager to make the final decision. So the model d the label and it won't need the manager to agree w The manager might want to know the both explana also think the manager should also provide the moc local explanation. I don't want the model to provide information I have known." Figure 7.

Feedback based on the last one

Only pay attention to the change

Activity - 2

■Participant 1 ■ Researcher

10. "Now use the message cards to order for the Scenario 2 which you makes." Figure 3. "I think it is very simple. I think there should be no feedback on the process because the manager has done it in the first time."
11. "What if the manager wants the model to give the both explanations again? " "What do you mean? " "I mean although the model now just shows a sample but it has learned something form the Scenario 1. So what do you think of it?" "Well I get it. In my case, I don't think the manager needs the explanations again because Scenario 1 the model has asked and the manager agrees with the final prediction or stuff like that." "So do you think of the model provides the last label in this time? " "It is hard to say. Maybe the real users could answer. At least for me, there is no effect if the applicant edits the CV or not. Only if I want to check if the applicant makes a fake CV in the second time."
12. "If the labels from the two workflows are different, do you need the both explanation?" "I prefer to ask the applicant immediately."
13. "Remember the first time the manager provides a lot of feedback for the model, right? So what do you think of the future model based on these feedback?" "Maybe. It likes the manager works for the model. To be honest, the model seems more to work for the manager but in the second scenario, the applicant and the model gets a lot from the manager. "

Transcription for Participant 2

One sentence includes 2 possible messages

Confused on 'feedback' & 'request to change' here

What & who gives the "scored" label?

Hard to see difference between "Sample and label" and "Sample for label"

Pre-test interview

P2
Role: A student majoring in Strategic Product Design (SPD) currently in TUD
Design/design education experience: About 5 years
Experienced in design & AI: One project about the branding strategic on one AI product.

Activity - Pre

■ Participant 2 ■ Researcher

Figure 8

1. "May I fill in multiple types of information in one card? " "Only one."
2. "Why do you think it is the same type of information in the 3th message and the 5th message? Can you tell the difference? " "Because both of these messages are asking for this model's opinion of the given CV. One is that the manager agrees with the view and the other is that the manager disagrees so asks for a change. Both are feedback against the prediction given by this model. " "What do you think of the 'edited label for prediction' and 'feedback for prediction' on the 5th message?" " 'edited label for prediction' is a final state while 'feedback for prediction' is a middle state. On the 5th message, I think it should be firstly a feedback and then an edited process. But I think there is only one I could fill in, I choose the feedback one. "
3. "Why do you think it is the 'Sample for label' in the 4th message and the 'Sample and label' in the 7th message? Can you tell the difference? " "Well, now I think they seem the same. Emm, the fourth message emphasizes the label "successful", while the "scored" expressed in the other does not necessarily mean successful, but rather what follows."
4. "Why do you think it is the same type of information in the final 2 messages? Can you tell the difference? " "Because decision is label, informing the same content, the object is different.."

Figure 8. Participant 2's pre

Model evaluates from the last feedbacks

Feedback based on the last one

Explanations will be asked for results that do not meet expectations, and vice versa.

Only pay attention to the change

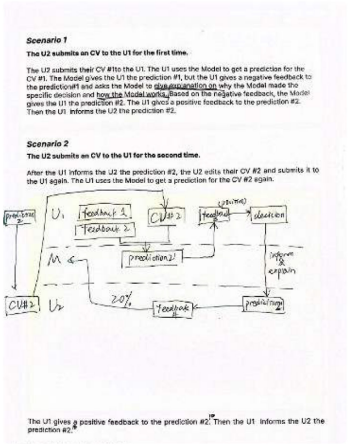
Activity - 1

■ Participant 2 ■ Researcher

Figure 9

1. "Why does the manager provide the Prediction #2 again in the second time? " "Forget it, it is a Prediction #2'. " "Oh, that makes sense."
2. "Can you explain why the story looks like this? " "When the model has to make a second judgment, it combines the first judgment with the feedback received to make a second judgment. I think this time the manager will give a positive feedback because the model has received a lot of feedback in the first time. If the model still makes mistakes in the second time, then I think the model might not work well. Then this time the manager will give the feedback directly and will not ask for explanations because the first time the manager has already asked for explanations of the model. And this time the feedback is positive and I don't think there is a need to seek further explanations. Then the applicant can be notified."

Figure 9. Participant 2's story



Activity - 2

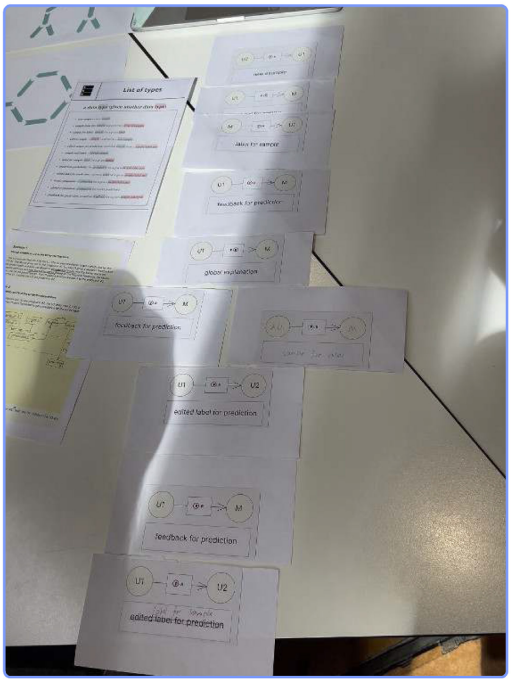


Figure 10. Sequence for the Scenario 1 of Participant 2

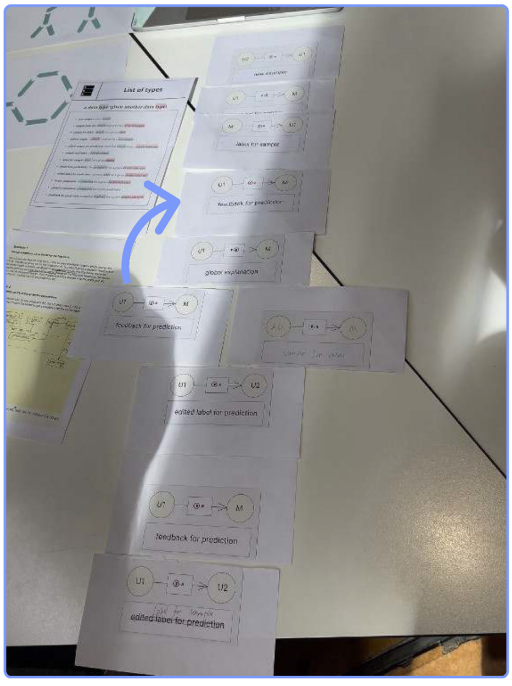


Figure 11. Loop on the Scenario 1 sequence

There will be internal loops, but in general, it's best to keep it short and quick

Model evaluates from the last feedbacks

Explanations from the actors to the model on labels, example, local & global explanations.

Explanations will be asked for results that do not meet expectations, and vice versa.

Model evaluates from the last feedbacks

Permissions between the actors and the model

Feedback from different actors for the model

Explanations will be asked for results that do not meet expectations, and vice versa.

Ask for feedback in moderation

Only pay attention to the change

The results of the model should match the expected results as closely as possible, otherwise actors will develop a sense of distrust or disappointment in the model

All actors could get the both explanations but the way to show the explanations might be different

The goal of effective workflow -> quick & short

Specific needs of users would affect the interface.

Activity - 2

■ Participant 2 ■ Researcher

1. "Would you please explain why the sequence makes sense? " "There should be one loop like Figure 11, I forget that while I make the sequence. It's like a logic diagram, with "Yes" or "No" going to the next step or continuing the loop."
2. "The applicant provided the CV to the manager and then the manager asked the model to give a label. Then the model provided the first result for the manager after analyzing the data from the CV. But since the first manager's feedback is negative, the manager wants the model to give an explanation and this is where local and global explanations come in. Then the manager gives feedback to the interpreters again. Then maybe this manager can provide a successful example to this model to make this model better for machine learning. Then this model may adjust its label. Then the manager gives positive feedback on the second result and gives the label or prediction to the applicant this time. "
3. "Is there any thing you want to keep for the next scenario? You have to do the same thing for the next scenario. " "I haven't thought about it... Wait! Just now I am considering if it would be better if the applicant also provides some feedback for the model. But I am not sure... In this case the applicant must want to be successful, so their feedback might not be objective. And the model must work for the manager or the company, right? So in my opinion, the applicant's opinion can be heard, but the percentage of this opinion may only be 20%."
4. "May I edit the message card?" "Yes, go ahead." "I think it would be convenient for the next sequence to say difference between the previous prediction and the new one." Figure 12.

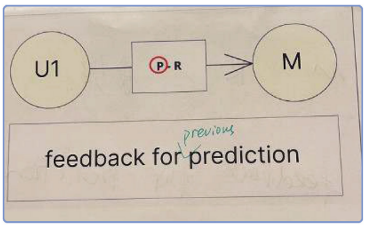


Figure 12. The edited message card

Activity - 2

■ Participant 2 ■ Researcher

Figure 13

1. "Oh I find I miss another part. It is about how the prediction #2' affects the edited CV. "
2. "Why do you think the sequence is like this?" "The applicant made CV revisions as suggested and then resubmitted. And this model predicts based on the feedback from the first time provided by the previous manager and this new CV, and the manager asks this model to give a new label. And based on the first time the manager helped the model to keep optimizing itself, so this time I expect the model not to keep needing feedback from the manager to optimize, but to do it all at once. So the manager gave positive feedback and notified the applicant. Finally, this applicant is giving feedback on this result, giving a weight of 20% to the model."
3. "If you work as a designer and has to evaluate the workflow, what would you do?" "What is it mean?" "To improve the user experience, the usability and so on. " "I think it can be seen by both the applicant and the manager when the manager asks for an explanation of the model. The equivalent of this system is an intermediate platform. This was my first idea, but then I felt that the feedback from the model would be rather cold, and it would be better if the manager could think about how to embellish the feedback or explanation from a 'human' point of view, so that the final message to the applicant would be more humanistic or 'warm' It may be better. "
4. "There is another scenario that is also worth thinking about. In the first time why this model gives negative results, it could be because the CV is not formatted correctly. This is something that can be fed back and modified. But if it's because this applicant's own ability doesn't match the position, then no amount of revision will meet the requirements, and these two programs may seem different to me."

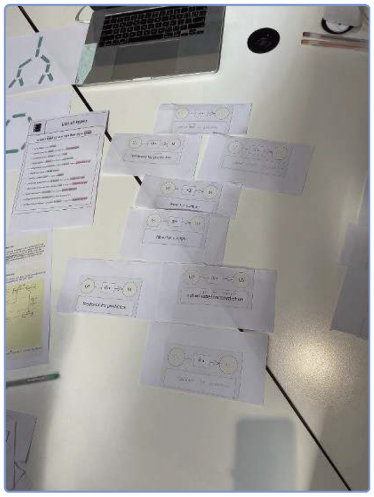


Figure 13. Sequence for the Scenario 2 of Participant 2

Activity - 2

■ Participant 2 ■ Researcher

5. "Do you think it would affect if the interface is different? " "Who is the user group?" "The manager or the applicant." "In my opinion, there are not so many chances for the applicant to communicate with the model. They might be more likely to communicate with the manager directly. If they want to talk with the model, the most possible questions are explanations. So I would prefer to give them on FAQ document. But for the manager, the chat box would be better. " "Do you think the style of the interface would affect the workflow?" "I don't think so. The logic for the whole story is the same."

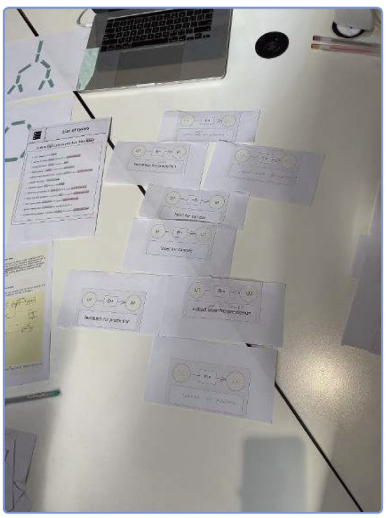


Figure 13. Sequence for the Scenario 2 of Participant 2

Possible decision–makers

This section simply records the relevant discussions on "who would make the final decision" during the project.

Figure 1 & 2 & 3 illustrate at least three distinct situations where decision rights differ between humans and the models. Participants asserted that clarifying these situations before constructing the interaction was imperative, as it could significantly impact subsequent interaction behaviors.

As shown in *Figure 1* , for the final decision, the feedback from the human and the feedback from the AI models are weighted 50% each.

As shown in *Figure 2* , for the final decision, the human can use feedback to keep the AI models working or modify the result until the human is satisfied.

As shown in *Figure 3* , for the final decision, the results of the AI models will not be influenced by human feedback and will only change due to changes in the input data.

“It depends on the how the manager and the model could affect the final result. While I am ordering the message, the model works like an assistant. That means it is the manager who makes the final decision." — Participant 1

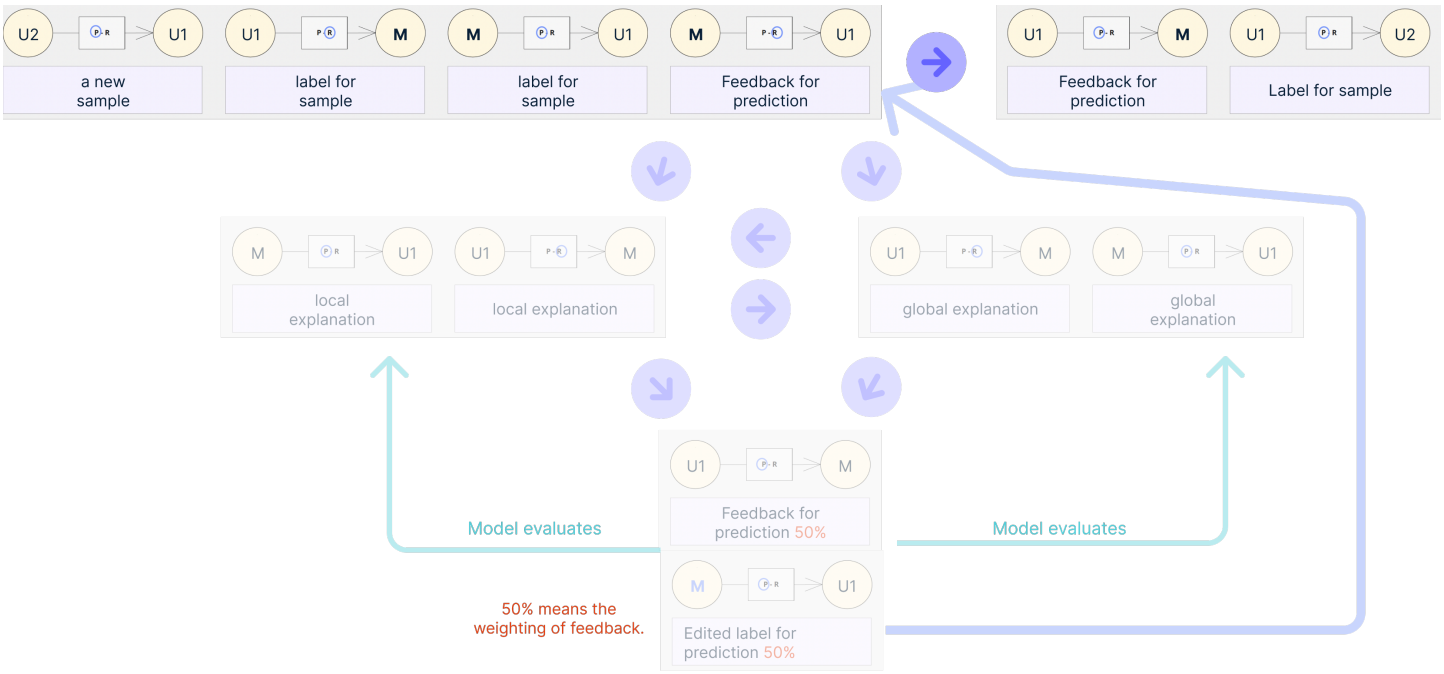


Figure 1. Both humans and the model have 50% of the decision–making power.

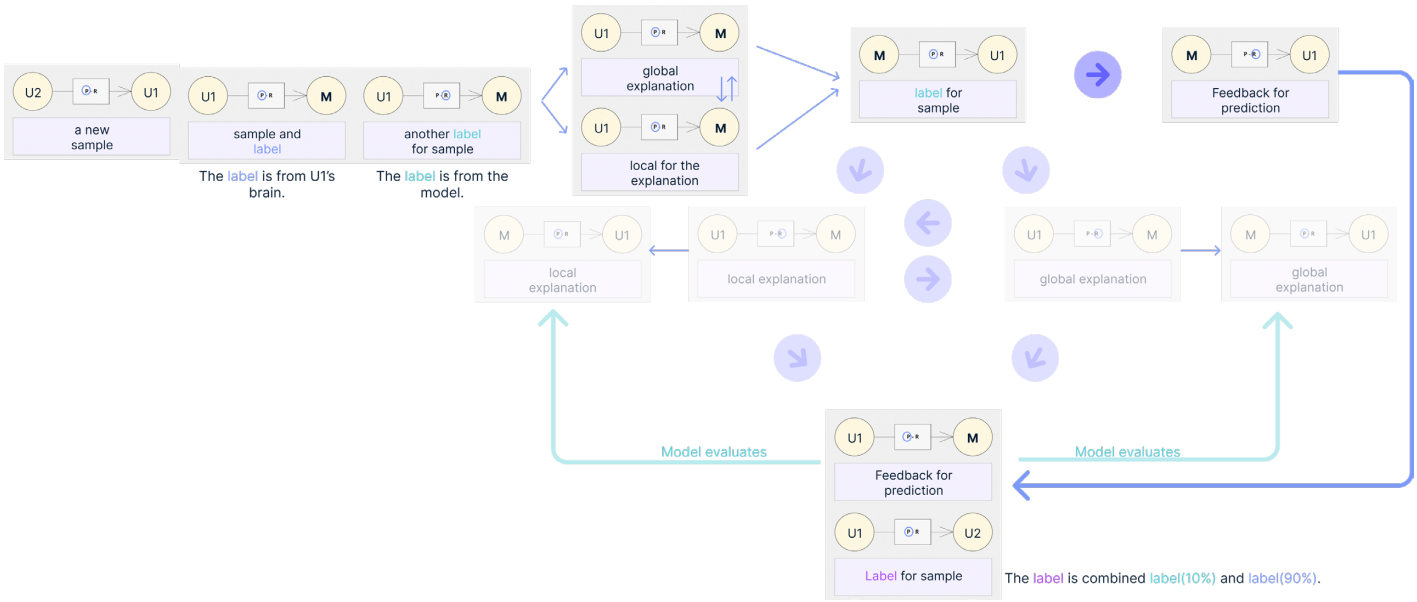


Figure 2. Humans have more than 50% of the decision–making power.

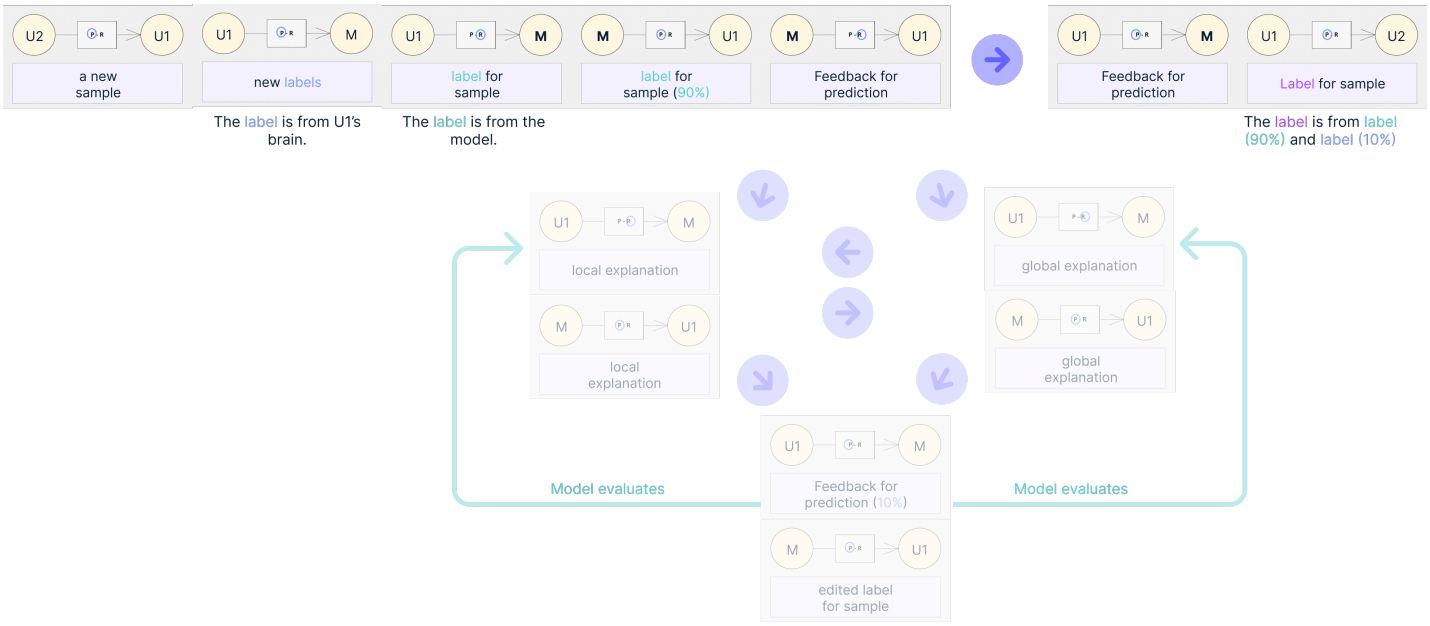
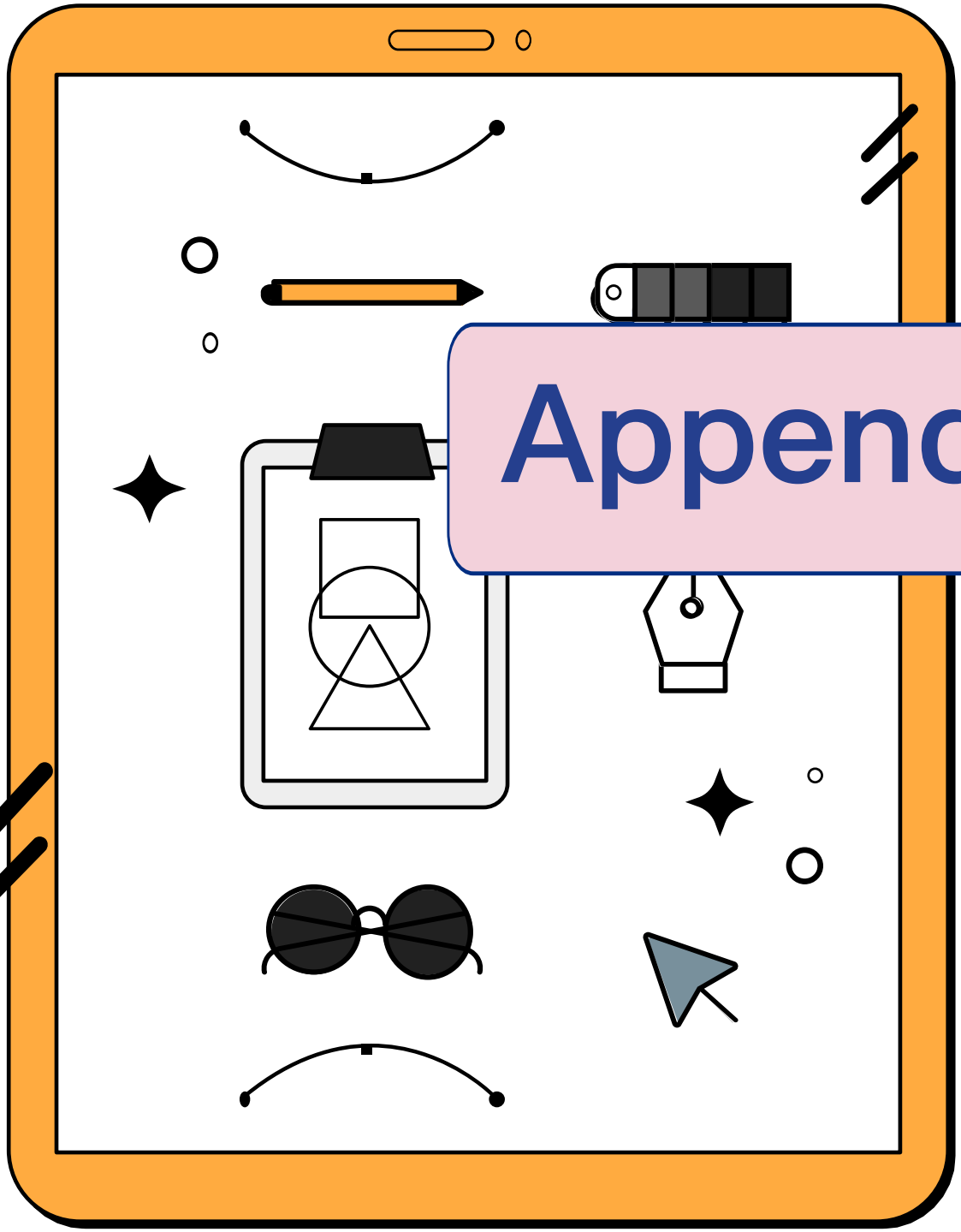


Figure 3. Models have more than 50% of the decision–making power.

Since this section was not one of the main purposes of this project, there was not much more exploration about how each type could be more accurately defined or how it could be used in a scenario.



Appendix F

Time: 21 August - 25 August

PROCEDURE
What are the main steps in the test procedure?

0 - 2 Min Welcome the participant	Introduce the test and ourselves. Ask user consent!
0 - 2 Min Check the prototype	Make sure that the prototpye is ready. Check if the recording works for tracking the test activity.
2 - 5 Min Pre test interview	General background questions to get a better picture of the participant.
20 - 30 min Work for Activity 0	During this part, show the participant how the prototype works and the meaning of concepts.
20 - 30 Min Work for Activity 1	During this part, show them the context. Then participants do Activity 1.
10 - 15 Min Interview	Interview about what the participant do in the 1st activity.

Total: 52 min - 84 min

LOCATION & DATES
Where and when will the test take place? When and how will the results be shared?

The test is conducted at IDE or Library in Delft. Each test will be conducted with one participant at a time.

EQUIPMENT
What equipment is required? How will you record the data?

- Tools to help record the test include, paper & pen, camera, computer and charger.
- Digital prototype in Figma.
- Camera for making pictures.
- Consent forms.
- Chairs and tables.
- Explanation materials for the context.

TEST TASKS
What are the test tasks?

Context: Designers work for designing interactions for the CV use case.

Activity 0: The activity includes 3 steps:
1. Introduce what the use case is;
2. Introduce what message cards mean, what the basic HAX means and what the meaning of the Instance & UI levels.
3. Introduce how to interact with the message cards and how the they connect with Instance & UI levels.

Activity 1: The activity includes 9 steps:
1. Ask participants to add any message card into the responding place;
2. Ask participants to connect each message card to information or interaction in Instance/UI level.

PARTICIPANTS
How many participants will be recruited? What are their key characteristics?

5 - 10 participants and they would be students majoring in design.

INTRODUCTION
"Hi, welcome to this test, I am Jiayi Zhou and thank you for participating in the test. The test and interview afterwards will be recorded, the video is only used for me and the project team analysis and will be deleted after the project is finished. By signing the consent form you will agree with this. The test consists of three parts: First I will ask you some questions to know your basic information. Then I will show you how the prototype works and the meaning of each concept using in the prototype. After that, you need to complete the Activity 2.If possible, it is better for you to think out loud while performing. And if there is any question, you could ask me for explanation or help. I would help you if the question doesn't affect the test. At the end of the test I will have a short interview with you. You could give me any feedback, positive or negative both are very important. Do you have any questions?"

PRE-TEST INFORM
What basic information the test needs?
1. Explanation on limitations of the low-fi prototype, "Message cards", "Human-AI interactions" and "Type of information".
2. Explanation on the context.
3. Fill in the Consent form.

PRE-TEST QUESTIONS
What basic information the test needs?
1. What's your name?
2. Are you a design student?
3. How long is your experience in design or getting an experience in design education?
4. Do you have experience on AI and design area?

QUESTIONS AFTER ACTIVITY 1
What questions would be asked after the test?
1. Is there any question you want to ask during the process? e.g. Something you don't understand?
2. Why do you think of adding this message?
3. Why do you think of adding this message here?
4. Why do you choose this kind of item in the Instance/UI level for the message?
5. How do you think of the dynamic model? Why do you want to show it in this way?
6. Is there any more suggestions?

Prototyping Human-AI Interactions - Activity 0

Objective: Assess if designers could understand instances of messages (label, sample, feedback), distribution of decision-making authority, and the iteration of models.

Materials
- user & model cards
- digital low-fi prototyping tool
- scenario card
- computer
- pen & white paper
- message cards & explanations

Task Description

Here is one Human-AI interaction. It shows how the applicant uploads CV, the manager asks the model to give a prediction of the CV, the model gives prediction and then the manager provides the feedback to the prediction.

Now the researcher would show you how the low-fi prototype would work.

Context:
#1 is one competitive applicant and good at 5 skills. However, the manager is not sure if the CV fits the job's description. So the model would give prediction. After analyzing the #1's CV, the model predicts it is accepted. The manager also agrees with the prediction.

Now that you are a designer who needs to build this resume judgment system, how do you think the interaction framework for the above process should be built based on user-friendly rules?

Steps:
- Choose the right message cards and order them
- Think about what instance would be shown under each message card.
- Think about what UI would be better to show the instances above.

Prototyping Human-AI Interactions - Activity 1

Objective: Assess if designers could understand instances of messages (label, sample, feedback), distribution of decision-making authority, and the iteration of models.

Materials
- user & model cards
- digital low-fi prototyping tool
- scenario card
- computer
- pen & white paper
- message cards & explanations

Task Description

Task 1:
After the Activity Pre, you may understand how the prototype works and the relationship between message, instance and UI.

Now the interaction is basic and you might feel it could be developed to be more user friendly. Go ahead! Add message you want to add and link it in the Instance or UI level.

Task 2:
The Task 1 shows one case if the manager agrees with the prediction. Now Task 2 shows one basic sequence on the manager disagrees with the prediction.

Context:
After analyzing the #1 CV, the model iterates itself and now there is one more new version called the #2 model.

#2 is one critical applicant. Although #2 has one highest skill, others are not so good. After analyzing the #2's CV, the manager is hesitate, so the manager uploads the CV ad ask the model to predict. However, the model predicts it is adjusted but the manager disagrees with the prediction.

Now that you are a designer who needs to build this resume judgment system, how do you think the interaction framework for the above process should be built based on user-friendly rules? P.s. you could create the #2 model freely and choose to use the model or the #2 model in the Task 2.

front

Instructions - Description of Materials

- Human-AI interactions are interactions between **human users** and **AI models**
- Users and models can exchange model-based **types of information**
- **User Cards:** describe (a type of) user based on their role in the interaction and their AI/domain expertise
- **Model Cards:** describe the model used in the interaction, including the model type and the model-specific types (input, output, feedback)

back

Instructions - Description of Materials

- **Types of information:** There are 13 types of information that can be exchanged between users and models
- A **message** describes the communication between a sender and a receiver in the context of a use case (*User and Model cards*). The sender can *provide* a type to the receiver or *request* a type *from the receiver*
- A sequence of **messages** can describe an interaction scenario between specific users and models.

Definitions and terms

- **Users** and **models** interact with each other using **messages**
- **Messages** are used to **provide** or **request** information
  - The sender (user or model) can provide information to the receiver (user or model)
  - The sender (user or model) can request information from the receiver (user or model)
- **Messages** use **communicative acts** to describe the **type of information:**
  - Communicative acts are defined based on what type of information a model can provide based on its specifications, i.e., model input/output, model metrics, explanations, and others.
- We provide a representation of such communicative acts using the cards

new sample

a new sample

- **sample: model input** an instance from input data, e.g., image, sound, form, video, etc.
- the new sample is in the model input format but it is probably new to the model
- users generate new samples by uploading a new file, capturing an image, etc.
- models can create a new sample, e.g., random feature values
- **Examples:** capture a new face, create a new document, create a new feature vector

sample from list

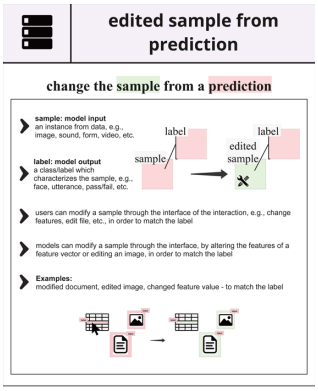
sample selected from a list of samples

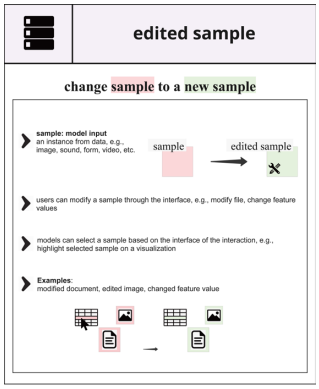
- **sample: model input** an instance from data, e.g., image, sound, form, video, etc.
- **list of samples: model input** a list/collection of samples
- user can select a sample through the interface of the interaction, e.g., click
- models can select a sample based on the interface of the interaction, e.g., highlight selected sample on a visualization
- **Examples:** select a file, choose a feature vector, click on a data point (visualization)

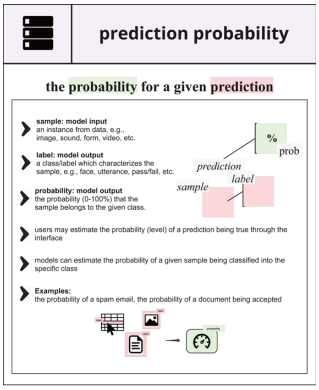
sample and label

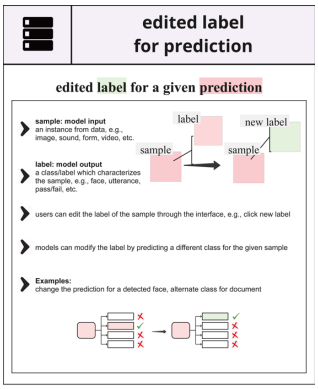
sample and label

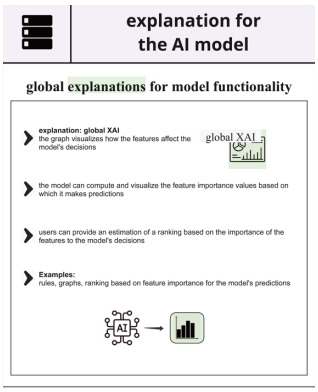
- **sample: model input** an instance from data, e.g., image, sound, form, video, etc.
- **label: model output** a class/label which characterizes the sample, e.g., face, utterance, pass/fail, etc.
- users can provide an example - a labeled sample through the interface; upload or select a sample with its label
- models can communicate a labeled sample through the interface, e.g., visualize a sample and its predicted/provided label
- **Examples:** annotated image, positive/negative example, new data point with label

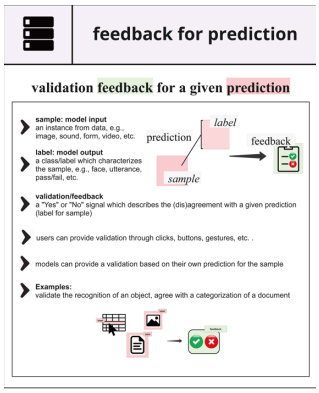


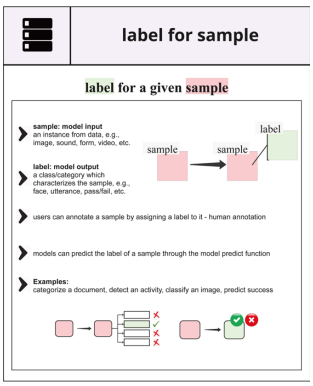


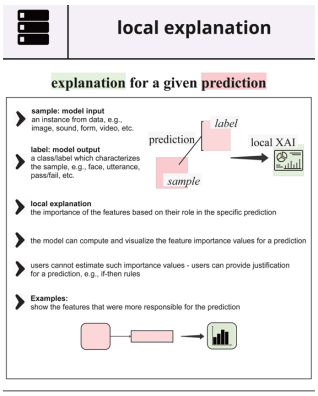


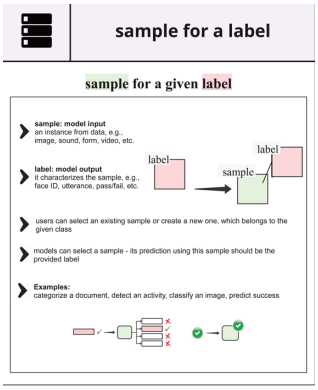




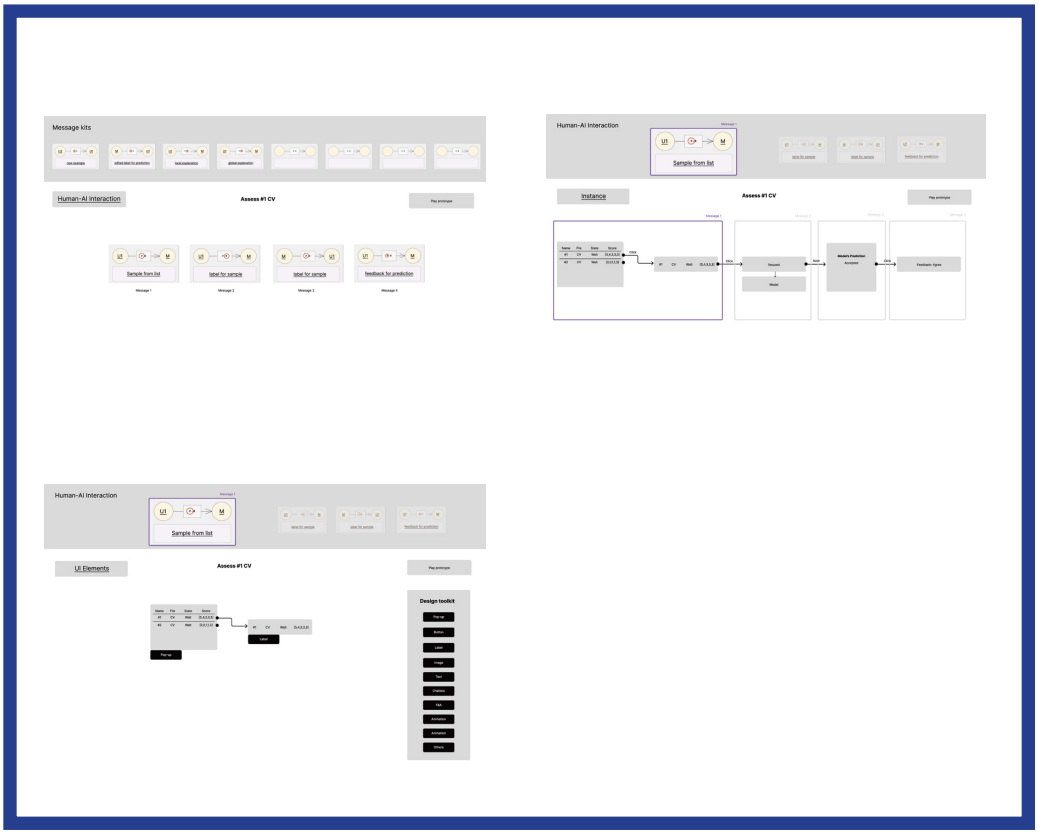








Interfaces for testing dueing Test 1



Appendix G

Transcription for Participant 1

How to let the explanation make sense all the time?

Sample/label/prediction/ explanation/feedback should be explained both in academy and context.

Maybe a reasonable example would help designers for understanding.

More clear explanations on what the feedback/label/prediction works for.

XXX gives feedback for ?

Basic information

Participant: #1
Role: Design student in TU Delft, majoring in Designing for Interaction.
Design (education) experience: More than 5 years
Experience in Design & AI: ITD

■ Participant ■ Researcher

Test

Activity 0

1. "How can we understand the sample, the label and the prediction? Please show me some examples instead of academic terms." "Sample is the item the human actor provides to the model. Label means to give it a label or an existed label. Prediction means the result the model provides to the human actor." "So what's the difference between the label and the prediction?" "They are almost the same for humans, but for the machine, they are a little different."
2. "What's the meaning of edited sample for prediction? Who could do that?" "Every human actor or the model could do all things here and request or provide all things to each other. For example, the manager think the sample-label pair is incorrect, so the manager changes the sample. "
3. "What's the difference between local explanation and the global explanation?" "Local explanation is for the specific case, the global explanation is for all."
4. "What does the feedback work for? What's the prediction made of?" "Prediction is the same as the label. The feedback is for the label." "So it doesn't include probability?" "Now it doesn't include. But if that makes more sense for you, go ahead."

Activity 1

1.1.1 "I am curious about why the scores would be the input instead of output? It seems the whole workflow is less effective. For me, I think the model would assess the CV by themselves and output these scores. The manager just inputs the requirements. Well maybe the manager just inputs the job description and the model assesses what the job wants. Then the model scores the #1 CV and how much it is connected to the job. If it is the manager scores the CV, that is too much work for the manager. And there is no motivation for the manager to use the model, right? " "Just forget it." "OK. First, I want to add one interaction on check if the manager gives enough data. If not, the model should provide the manger with places to fill data in. " See Figure 1.1-1.2.

1.1.2 "I think I need to know the motivation for the manager to use the model. You see, if the manager just uses the model to assess 100 or more CVs and then get the top 5%, it needs one interaction flow, However, if the manager wants the model to compare #1 and #2, there are another interactions. " "You could show me these two different workflows. "

1.1.3 "Well, does the feedback would affect the model? You talk about the dynamic characters, right?" "Maybe. If that makes sense." "It could be different from the feedback for the prediction, I think. It is a little ridiculous if the model iterates itself during the process. Of course, it would be better and better but how about the past results? I mean, in the beginning, it doesn't work so well so it might lose some good CVs. That might make the manager doesn't trust the model. So I would add one more feedback special for the model iteration. " See Figure 2.1-2.2.

The context is more reasonable, the better participants would work.

Freedom to edit the input & output

Different motivation, different message card orders.

Flexible ways to edit sequences

More clear explanations on what the feedback/label/prediction works for.

Models work with an improved approach (sequestration/dynamic)

Confidence on the model's results

Different feedback for different goals (model -> model, result -> result)

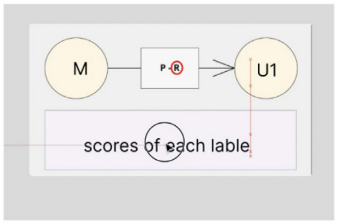


Figure 1.1. Confirm 1

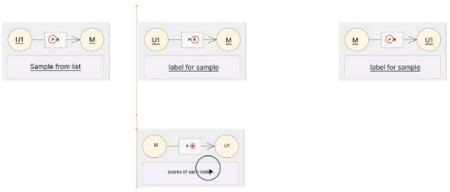


Figure 1.2. Confirm 1's place

How do we define the Confirmation information? Feedback?

Or directly jump into the page to refill the lost information.

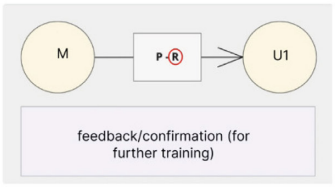


Figure 2.1. Feedback 1

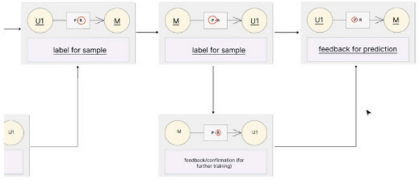


Figure 2.2. Feedback 1's place

Edit the prediction freely. (only label/label + probability/ only probability)

Show how the model changes to the manager.

Can do/cannot do/how to do

Abilities of different roles(human actors/model)

Think about how we could help designers to create their own message cards/type of information

Many situations we might not presee.

How do we make sure the designers use message cards right?

1.1.4 "I think in this message card, it would be better to provide more information like this." See Figure 3.

1.1.5 "If the manager gives a positive feedback for the prediction, then what would be affected or changed? " "It would affect firstly the label provides for the applicant. Secondly, it would work for the model's iteration." "And I think the iteration information is a little like model-to-model information, right?" "Why?" "It tells itself to think more like the manager. " "Ok. So do you think the manager should know that? " "Yes, of course." See Figure 4.

1.1.6 "If the manager doesn't agree with the prediction, does the model edit the label or the manager edit that?" "I think it should be the model to edit it," "So it should be the U1 request the model." See Figure 5.

1.1.7 "So the first time, the result shows no. Then what would happen? I mean does the applicant have the second chance to apply for the job again?" "For this time, we think the applicant cannot do that. For a new job, yes. Is there any difference between these times?" "Well, if the time period is not so long, you know, the label shouldn't change a lot. Of course there would be people who develop their skills in a short time, but just forget them. To be honest, I have one more question here. If the applicant uploads the edited sample in a short time, is the CV a new sample or an edited sample? " "Good question. For me, I think it is the new sample because it applies for other jobs. " "If it is for the same job?" "I need to check." "Or maybe there would be one place to store all of these cvs."

1.1.8 "Now for the Instance level. Is it data which the message card includes, right?" "Yes." "Well, is it possible to change the the list? " "Of course, everything you could change." "Including CV, is there any other type of files? For example, portfolio? Because I think it is a little strange if the CV has to compare with other files like portfolios, I hope there is only one type of file here." "Only CV." "In that way, I don't think we need the file here. #1 means the #1 CV. " "Great." See Figure 6. "I want to create instance on the Message level interface. Fewer interfaces, more effect."

1.1.9 See Figure 7. "I think it should be the model's behavior instead of the Request because it is the model who gives the prediction."

1.1.10 See Figure 8. "I think firstly, the manager should input data like scores, requirements and then they get the list. And before that, the model may ask the manager to provide it."

1.1.11 See Figure 9. "I am thinking about the way the manager disagrees with the prediction. Does the manager have to provide reasons why disagreement? " "How do you think of it?" "It depends on the manager. If the reason is acceptable or reasonable, for example, the applicant is the best at one skill while others are not so good, It is reasonable and the manager could write down, However, if the manager just thinks the applicant is pretty. It is not reasonable and something cannot be write down." "How about asking the manager to give reasons all the time?" "No, I don't think it would be a good idea, because the reason doesn't exist for the all time. And if the manager always has to provide the reason, it would bother them and maybe they write nothing all the time. "

1.1.12 "After the disagreement, the model should edit the label directly. Well, I think the decision situation is a little complex. "

Cannot design one element without other elements (or the background)

How do designers could merger all elements/instance together?

Maybe it is a funcition part?

Appropriate frequency or manner of asking for reasons

Ability not to answer questions they do not want

Disagreement situation is something we need to think about more.

Decision part should be chosen before the final step.

Transcription for Participant 1

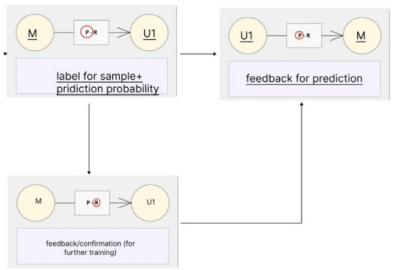


Figure 3. More information

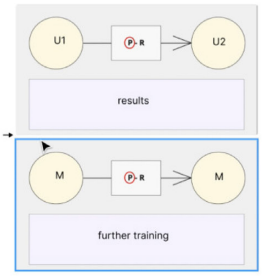


Figure 4. Training 1

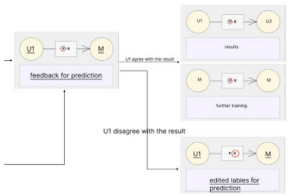


Figure 5. Disagreement

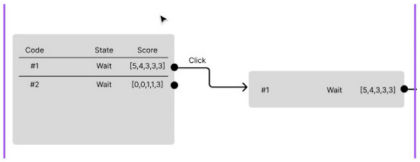


Figure 6. Effect information

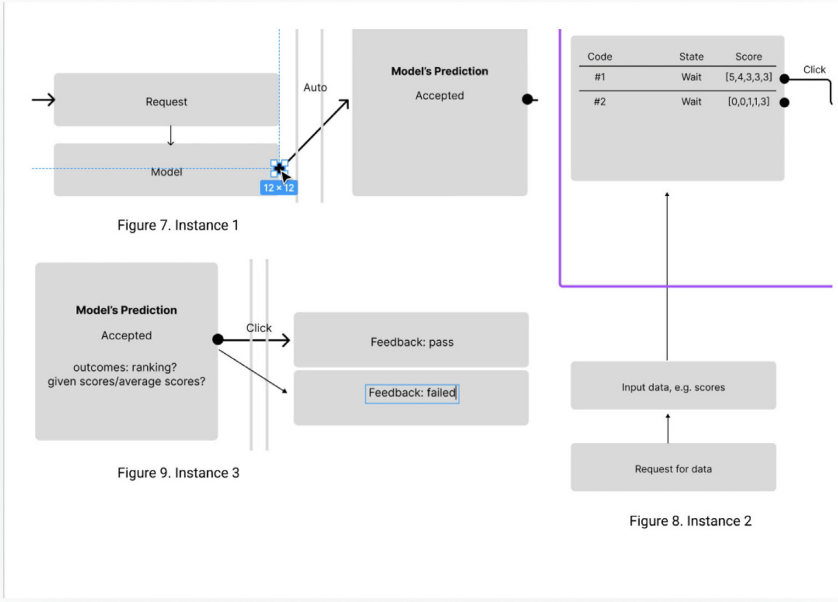


Figure 7. Instance 1

Figure 9. Instance 3

Figure 8. Instance 2

Prediction should be edited by designers based on context.

Change the model based on the context/needs.

Situations might be more complex if there are multiple labels.

One way to renew the label.

Easy to edit U elements on the instances

Related to other design elements easily

Sequestration before practising

1.1.13 “The final steps are complex. The prediction is only two answers, yes or no. No scores and other things. Then to be honest, I don’t see any effect the manager’s feedback has. I mean if the prediction is yes, the manager disagrees with it. Then that means it must be no. Why does the machine work? ”

1.1.14 See Figure 9. “Maybe we could change a little to make it reasonable. You see, the prediction is to give a ranking on outcome, then the manager could choose pass or not pass. ”

1.1.15 See Figure 10. “I also think maybe there would be one way to go back. Like the doubt. Then the model needs to assess the CV again.”

1.1.16 “For the UI Level, I think it would be convenient to edit it in the Instance level. It just needs one label, right? And it is a web interface or an applicant?” “Web interfaces.”

1.1.17 “I think now the workflow is only for human-AI interactions. For the real case, I would start from the homepage. And I think the manager should deal with a lot of samples at the same time. And then why do we have to select and assess the #1? ” “So again, based on the manager’s needs.” “Yes.”

1.1.18 “For the reliable results, I hope it could be practice before using. ”

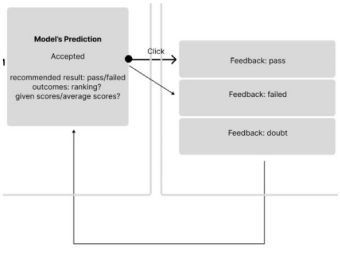


Figure 10. Instance 4

Transcription for Participant 2

Abilities for human actors & models

Use examples related to the context to make them understandable.

Q&A for some common questions (two explanations, edited samples ...)

Basic information

Participant: #2
Role: Design student in TU Delft, majoring in Integrated Product Design.
Design (education) experience: 2 years
Experience in Design & AI: Graduation project about Chatgpt

■ Participant ■ Researcher

Test

Activity 0

1. "Who could pass these information?" "All human actors or the model could request or provide any type of information to others." "Is there any limitation?" "If it makes sense for you, go ahead."
2. "What's the difference between local explanation and global explanation?" "Every sample has its own local explanation, and all samples have the same global explanation. Global explanation works like the basic rule for the model."
3. "Is the feedback for the prediction or the probability?" "It depends on you."
4. "What would happen if the human user uploads one edited sample? Does the model still give the same label? Is the edited sample a new sample? How does the model could know that?"

Activity 1

1.1.1 "I would pay more attention to the way the model changes itself based on the feedback from the model."

1.1.2 See Figure 11. "I think after the whole workflow, the model could ask the feedback on the whole workflow." "So do you think the steps appear when the manager agrees with the prediction?" "I think these steps appear when the manager disagrees with the prediction. And I don't think there's anything in there that accurately describes this type of information. "

1.1.3 "For the instance level, although the task says the manager would agree with the prediction, I think we still need to provide all options for them. " See Figure 12.

1.1.4 See Figure 13. "I think the model could ask the manager to give explanations and suggestions on how to work better. For example, show the 5 input aspects and scores, then ask the manager to choose which one should be developed. I hope to define them before message levels. "

1.1.5 See Figure 14. "Maybe we could show them five buttons and then they could click and write details. That is more about the UI elements."

1.1.6 "Is it necessary to store what the manager's feedback is?" "Of course! If not, the manager has to do the same again and again. And the model should iterate itself to meet the manager's needs. "

1.1.7 "I think the list is one interface, because it is important and there are a lot of CVs here. For the start button and the prediction, I prefer the pop-up. For the feedback, I hope they would be buttons. When people have to think about the exact choice for developing. I would prefer the buttons on the interface." See Figure 15.

The way the model iterates itself is one important thing.

Different feedback for different goals.

All choices should be shown even if the manager might not use.

Easy and fast ways for uses to set before the model works.

Easy and fast ways for uses to set before the model works.

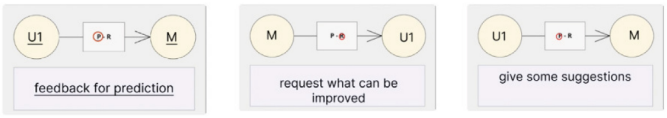


Figure 11. Model 1

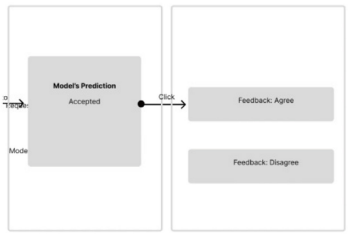


Figure 12. Instance 1

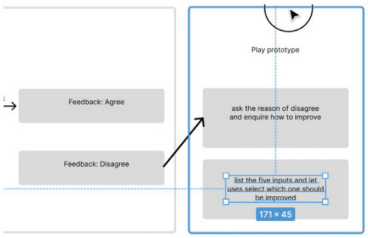


Figure 13. Instance 2

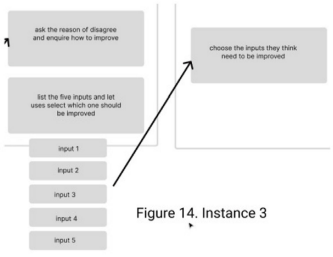


Figure 14. Instance 3

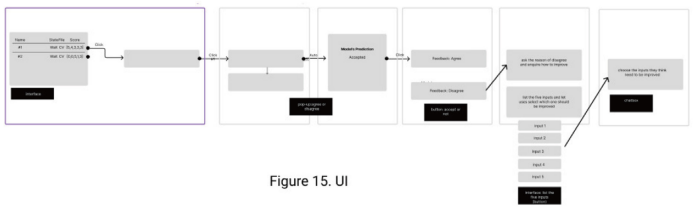


Figure 15. UI

Transcription for Participant 2

Chatbox is easily to input text but sometimes it wastes time.

Enough triggers for information pass.

1.1.7 “When the manager has to provide the suggestions or explanations in detail, I prefer a chat box.” See Figure 15.

1.1.8 “If the manager has to get top5 of 100CVs, is there any difference?” “Of course. I think there would be one interface to show all CVs and then there is one button for filter. Then go to the current message flow. So I think the first is a new list. ” See Figure 16.

1.1.9 See Figure 17. “So how about the instance and UI elements?” “The first place is filter, it is a button and the manager needs to choose the requirements. For example, how to order the top 5? Then we need to think about how to show the top 5. For me, the interface is only show the top 5 instead of all CVs again.”

1.1.10 “What’s more, I think it would be better if we could set one standard to assess the scores. For example, 5 is the lowest score. If the CV has more scores than 5, that means the CV is successful.”

1.1.11“I think Message is too abstract. When I need to create connection and UI components, I feel relaxed.”

Figure 16. Message 2

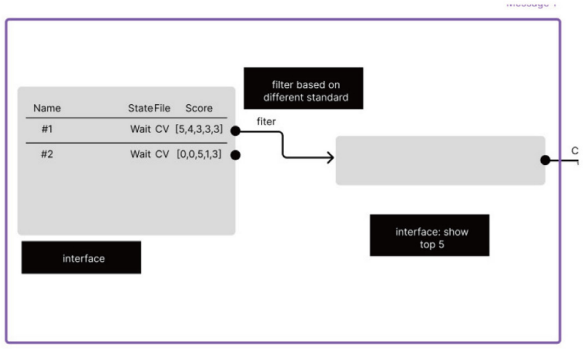


Figure 17. Instance 4

Transcription for Participant 3

Suitable way to visual the meaning.

Show meaning of every term before testing/using.

Examples of explanations related to the context.

Abilities to edit the message card/terms/...

Basic information

Participant: #3
Role: Design student in TU Delft, majoring in Strategic Product Design.
Design (education) experience: 5 years
Experience in Design & AI: None

■ Participant ■ Researcher

Test

Activity 0

1. "What's the meaning of sample? Could you show me examples about it? And if possible, also the meaning of prediction, label, probability, explanation and feedback." "To make it easier, I would introduce the context for the later task. Sample is the CV given the model, and prediction is the same as label. Label means the result of assessing, in the context, pass or fail. Explanation means reasons why the model gives the prediction. Feedback means the reflection on the prediction, or modify the prediction. " "So in the picture, the model needs to give feedback by itself?" "In most cases, it is the manager do. However, if you think that makes sense for you, go ahead. "
2. "In what kind of situation, someone needs to edited sample for prediction? And at the same time, when do we need edited label for prediction? " "For me, maybe when you use the label to find the responding CV, then you disagree with the result, maybe you need to edited sample for prediction. " "Ok, make sense."
3. "What's the meaning of sample-label pair?" "That means the sample has its own label. " "But why do we edited them? Or when could we do that? " "You think it is not reasonable for the pair. " "So we could change the pair, right? " "Yes."
4. "Why do we edit sample? " "Maybe the model cannot read the last CV style so the applicant uploads the new one." "Make sense."

Activity 1

1.1.1 "So when the designers create the interaction way for the system, are the labels sure? I mean, we have known what kind of labels we need." "Well, we don't know."

1.1.2 "Why there is no message card or type of information talking about the model gives prediction to the human actor?" "Good question. I always use the Label for sample to show it. Remember, the label is the same as the prediction. " "So why there are two different words representing the same meaning?"

1.1.3 See Figure 18. "Here is the best sequence for me. May I explain it? For the first message card, I think the start would be the manager provides the model what kind of labels he needs. Although there are scores for these five aspects, I don't know if the model know them or not. We need to let the model understand these numbers, right? Then the manager could provide the #1's CV to the model. After that, the manager would ask the model to label for sample. And the model provides the label for sample. Here, there are many optional choices. For me, maybe the manager would edit the label to assess or add some new labels. If the manager edits labels, we also need the scores. So where could we upload the score? Then, the manager requests the model to provide the prediction again. Seeing the prediction, the manager may need local explanation and then give feedback. Of course, for me, maybe the manager doesn't agree with the model, and may edit the setting of the model."

Use the same words for the same content

Edit the model's input/output

The way to set what human actors want in the model.

Abilities on human actors & models

Explanations are only shown if necessary

The content of the prediction would affect the interactions

Trigger is something not in the instance level.

Easy way to get the goal

More accurately to enhance

We need to provide the manager with places to write reasons/explanations

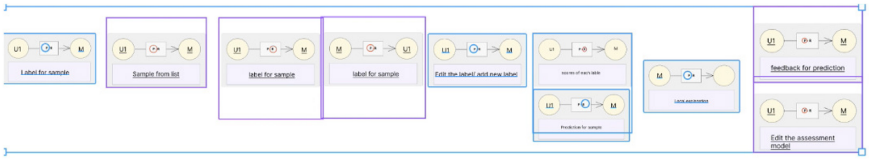


Figure 18. Message 1

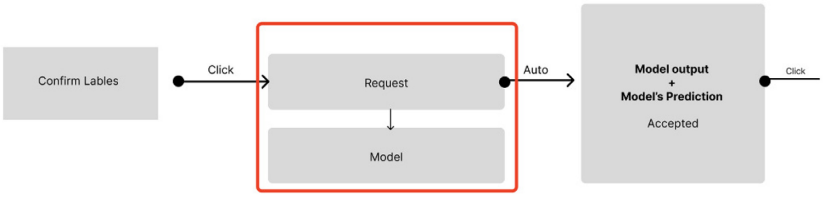


Figure 19. Instance 1

Sometimes, designers need to change their interactions.

Prepared materials could be provided to make the process fast & easy.

The more reasonable the context is, the better designers could work.

UI elements are always related to other elements in the same interface.

Explanations are needed mostly when the manager disagrees. The frequency would affect the user experience.

1.1.4 "Wait! I want to change a little on my explanation. For the third and the forth steps, maybe there are some prepared labels, the manager could choose or change these labels based on the special context."

1.1.5 See Figure 19. "What's the meaning of the Request-Model part?" "It means there is one trigger to turn on the model."

1.1.6 "May I change a little on the message card sequence here? " "OF course."

1.1.7 "To be honest, I think the context is a little confusing. For me, it would be more reasonable if it is the model chooses the CV from the list. I cannot understand why it is the manager who picks up the CV. "

1.1.8 See Figure 21. "Here are the new message card sequence. For the first step, because after thinking it twice, I think if the manager knows who would be the right CV to assess, maybe the manager also knows what kind of skills are suitable for the job. So the manager could edit the labels directly. "

1.1.9 See Figure 22. "So for these two steps, their instances are like these. Firstly, we need one list, then the chosen CV. After that, text what the manager needs on the screen. Then check if these labels are what the manager wants. After the model works, the prediction shows the prediction and the probability. At that time, I was not sure if the local explanation in the message card should be shown with the prediction or not. Finally, I think maybe sometimes the manager doesn't need it. In a short word, the local explanation is not something we need for all the time. If the manager needs it, request for it. And for feedback, if everything is ok, just agree. However, "

"If the manager doesn't agree with the prediction, we have two situations, One is the manager needs to edit the model. The other one is the model needs to renew. "

1.1.10 "So do you think there would be difference for the model if the manager agrees or disagrees with the model's prediction?" "for me, I think we need to record why the manager disagrees with the prediction. Then the AI engineers or designers need to iterate the model based on reasons. And I am not sure if the designers could see and understand local and global explanations. IF yes, maybe the designers could edit them and make it more acceptable for managers. And maybe change the working rules for the model. "

1.1.11" For me, there would be a lot of CVs in the list. Then the manager could input words or other things as filters for the model to find CVs writing them. Then the model gives these CVs scores and makes an order. After that, the model shows the overview of these CVs. The manager compares two or three CVs. "

1.1.12 "For the final result, I think it needs both of them, For onr thing, the model shows the data. For the other thing, the manager could write some more good words to let the results acceptable."

Make the feedback clear. What would be affected by feedback?

Reasons why disagree.

Designers could work for the local explanations/global explanations.

More related to the context

Right for desingers to change something in the model.

Maybe the "decision" part is something not so complex. We could get benefit from both the model and the manager.

Transcription for Participant 3

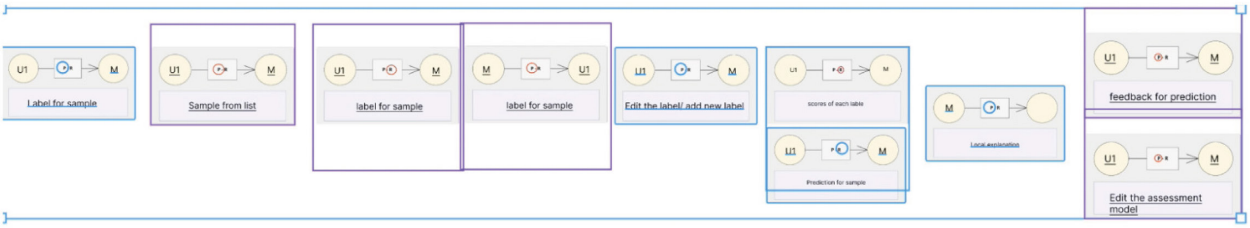


Figure 20. Instance 2

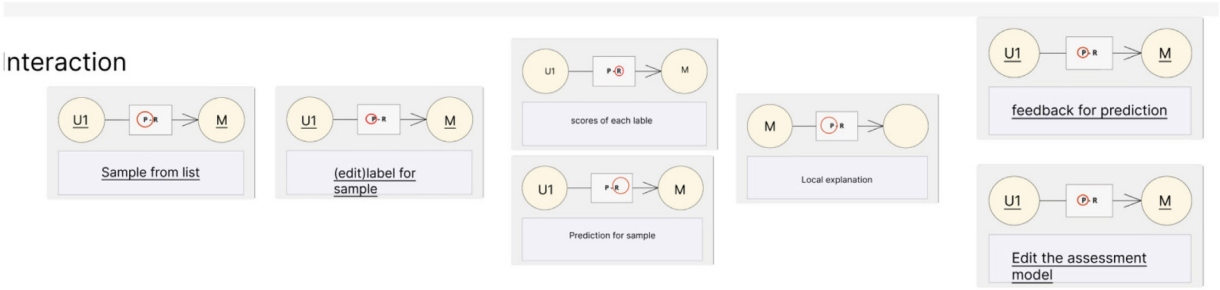


Figure 21. Message 2

Figure 22. Instance 3

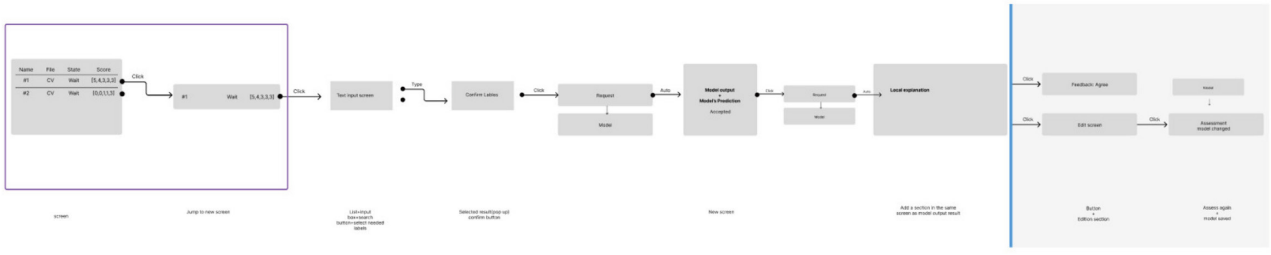


Figure 23. UI 1

Transcription for Participant 4&5

How to show difference in some terms?

Abilities for human actors & models to edit the message card/terms

What does the feedback work for?

Basic information

Participant: #4
Role: Design student in TU Delft, majoring in Design for Interaction.
Design (education) experience: 5 years
Experience in Design & AI: Working experience in-car AI assistant (about 1 year)

■ Participant ■ Researcher

Test

Activity 0

1. "What's the meaning of local & global explanation? " "Every specific sample has its own local explanation. All specific samples have the same global explanation. "
2. "What's the meaning of the prediction probability? Is it the probability for the sample-label pair or for the successful rate?" "I think they are the same." "Maybe they are a little different for me. If the probability is for the sample-label, that means the sample-label pair is correct. And this means that the likelihood of this one pair appearing. However, if the probability is for the successful rate, that means the sample-label is not confirmed and maybe it is wrong. " "For this context, it is the first choice."
3. "Who could edit label for prediction? " "All human actors or the model could do that if it makes sense for you."
4. "What's the feedback for? " "The prediction is the same as the label. So in this context, it could be the feedback to the label. " "Also for the probability?" "If that works for you, yes."

Basic information

Participant: #5
Role: Design student in TU Delft, majoring in Design for Interaction.
Design (education) experience: 3 years
Experience in Design & AI: None

■ Participant ■ Researcher

Test

Activity 0

1. "I am curious about the edited sample and the new sample. If the sample changes, is it a new sample or is it an edited sample? " "It should be the edited sample. But you could see the explanation, in somehow it is also a new sample." "So how does the model work if there is one edited sample? Does it give the same label for it or it would change? " "It depends on you."

How to show difference in some terms?

■ Participant 4 ■ Participant 5 ■ Researcher

Activity 1

1.1.1 "I had some thoughts before making adjustments to this message order. I think maybe this message card doesn't necessarily need to be in the form of a card, maybe as a tab would be a good way to go." "How about physical cards?" "Tab would be better for digital tools. Then you could have four tabs, and in each tab, you could edit the data, information or anything you want. Finally, you could play the whole prototype together."

1.1.2 "Now we could see the final prototype is made of 4 message cards with ui and data. It is a little hard to understand. If we could see different tabs, maybe it would be easy to understand."

1.1.3 "I image the message card could be like the tab in Figma, then we edit everything we want in the tab freely. And we could merge all tabs in one place."

1.1.4 "So is it an application or?" "It could be the website, right?" "Yes, just imagine it is the website interface."

1.1.5 "To be honest, I think what you do is more like one function component in the design tool instead of the whole design tool. " "Now we pay attention to this function." "So can we use the model like the voice assistant? " "If you think it makes sense in the context, go ahead."

1.1.6 "I am thinking about the context. When the manager gets the system, he has to click buttons and then there is on list on the interface, right?" "It depends on you." "So the goal is to help designers with tools to build system interfaces for the context." "Yes."

1.1.7 "So how about adding some AI tools into the platform to help designers doing some design work. For example, show me some templates. Then I might set some templates for the manager, or guide the manager to set their own templates."

1.1.8 "I find the input is from the manager and it includes scores. That means the manager still has to assess the CV. In that case, I don't see any needs for using the model. I think maybe the manager could provide the job description to the model, then it is the model to decide what skills the job needs and find the correct CV. " "I agree. The scores from the manager are too subjective. For example ,the first CV seems so good and the manager gives it a high score. However, after reading all CVs, the first CV is the worst one. Maybe we need one standard to assess, not only for the manager, but also for the model. "

1.1.9 "I also think about maybe we could use different colors to highlight different words and labels. Then the manager could click different colors to learn more. So maybe the input would be better to be the words" "But take care! The words on the CV sometimes are confusing. It is a little hard for the model to see some difference or word games."

1.1.10 "You talk about different message card sequences in two situations. I wonder what would be the difference?" "Well, if there are a lot of CVs and all of them have scores, I would prefer a list which order all CVs in special limitations. If there is one special CV, it is the current order." "And if there are a lot of CVs, I think it would be better if the interface only shows the CVs meet requirement instead of the all. For example, only show the passed CVs."

1.1.11 "And for the prediction, I think it would be better if the prediction includes the probability of labels. If there is only one prediction, it is too sure, which may make the manager doubt."

1.1.12 "I think in the first interface, the list holds overviews of CVs. After clicking, we could see the details."

Templates may help designers learn how to use the interface.

Fast & easy way to finish the task.

Different UI functions

Limitations on what function we want to add into the digital tool?

Decision part may depend on the most effective work way.

If the model cannot work, the manager could do it.

Free to edit the prediction by designers.

Good way for designers to understand what they are doing and what would be the result.

How does the designers could see the working prototype?

How does the AI design component connect to the common design process?

Motivations are different in different design context.

1.1.13 "So how do you input or get the output?" "For me to input, I think it would be better to select the value from a range. And for scores, I prefer to give the manager right to set the standard score. Above the score, pass; below the score, fail. " "I think there would be two ways. First, the manager sets one basic score and the requirement and then inputs them into the model. The model generates possible words which might appear. The UI would look like the chatbox.Then the manager inputs the CV into the model. The model searches if the CV has these possible words. Second, the model gets the CV directly and assesses it. I think the second way might work worse than the first one."

1.1.14 "For AI, I think most people would prefer buttons or other interaction ways they could get feedback immediately instead of voice assistants. That is because people have to wait longer while using voice assistants. It is only be used when people cannot use hands at all. " "Buttons are more direct. And maybe some numeric bars." "Maybe for the "

1.1.15 " For the feedback, do you have any ideas?" "Well. it reminds me of shopping in the supermarket. For example, after you use the system, it asks you to give a score. It may work but it shouldn't appear frequently. I think in most cases, people don't care about it unless it has problems." "In most cases, if we are satisfied, we miss it. If it doesn't meet my needs, I would ask for the explanation and give a feedback. So the interaction would appear in people who are not satisfied."

1.1.16 "For local explanation, I think maybe it should also appear in the output. So the prediction is made of a text. Then it could also show the work rules. "

1.1.17 "When there are two same CVs, maybe there is one feedback after the comparing."

1.1.18 "Is it possible that one CV is great but the model predicts no. So what would happen?" "I think the manager might miss it. So maybe for the first use, it is better to check the model. " "The feedback for every CV seems necessary. "

1.1.19 "For special context, maybe the model could be teach by users before using."

1.1.20 "Maybe it is a little like the application does for users now. Just imagine it is a junior hr employee!"

1.1.21 "For different roles of the model, I think they are different. And we have to take care, the steps before the model to work shouldn't be too many. It is still preparation, and it makes the work pace too slow. "

Motivations are different in different design context.

Easy interactions & effective workflow

Appropriate frequency of explanation/feedback

The local explanation is a little different from the global explanation. Maybe human actors need it more.

Confidence on the model's results.

sequestration/dynamic

Transcription for Participant 6

Basic information

Participant: #6
Role: Design student in TU Delft, majoring in Integrated Product Design.
Design (education) experience: 2 years
Experience in Design & AI: None

■ Participant ■ Researcher

Test

Activity 0

1. "What's the probability mean? Does it work for the right rate or some other things?" "I think they are the same. For example, 80%Yes, either means that there is an 80% probability that the prediction Yes holds, or that there is an 80% probability that the prediction is correct. They all mean the same thing. "
2. "What's the local explanation or global explanation's meaning?" "The local explanation means the specific reasons on why it is the prediction. Every sample has its own explanation. The global one means basic working rules for the model. All samples would get the same explanation."
3. "How does the human user give the feedback? By writing or by clicking?" "It depends on you." "What would the result be? I mean the prediction, how does it look like?" "It also depends on you."
4. "What if "

Activity 1 - Task 1

1.1.1 "I am curious the stop & switch. Now they are the same page, but I don't know why. What does stop mean? Stop the model at all or stop the process? "

1.1.2 The way to show how the model #2 comes.

1.1.3 Where do I could see all types of information? I think they are one of the most important parts in the prototype but there is nothing about them now in these interfaces. Shall I read the paper material and then fill in the digital message?

1.1.4 To be honest, I don't have a lot of suggestions on the current workflow. It is the direct way to solve all problems, right? I like these kinds of direct way. If I have to say something about it, it must be the terms about types of information. It is a little hard to understand them and imagine what they are. You see, in most cases, you would give me some examples to get the definition of each term, like the CVs are samples in the system. What's more, the whole use case is a little strange. It doesn't make sense for me.

Activity 1 - Task 2

1.1.5 Now I have to add something about UI or interactions. By the way, I thought there was something strange about the whole process. Earlier we used the abstraction you provided to design these processes at a more generalized level. But now we are back to the more figurative concept. This seems to imply that earlier we spent some time doing useless work.

1.1.6 I think the design of the ui can have some comprehension difficulties as well. I don't understand use case, which leads to having some content that I don't know how to design. One CV and many CVs, are different.

More clear explanations on what the feedback/label/prediction works for.

Type of information is one important element

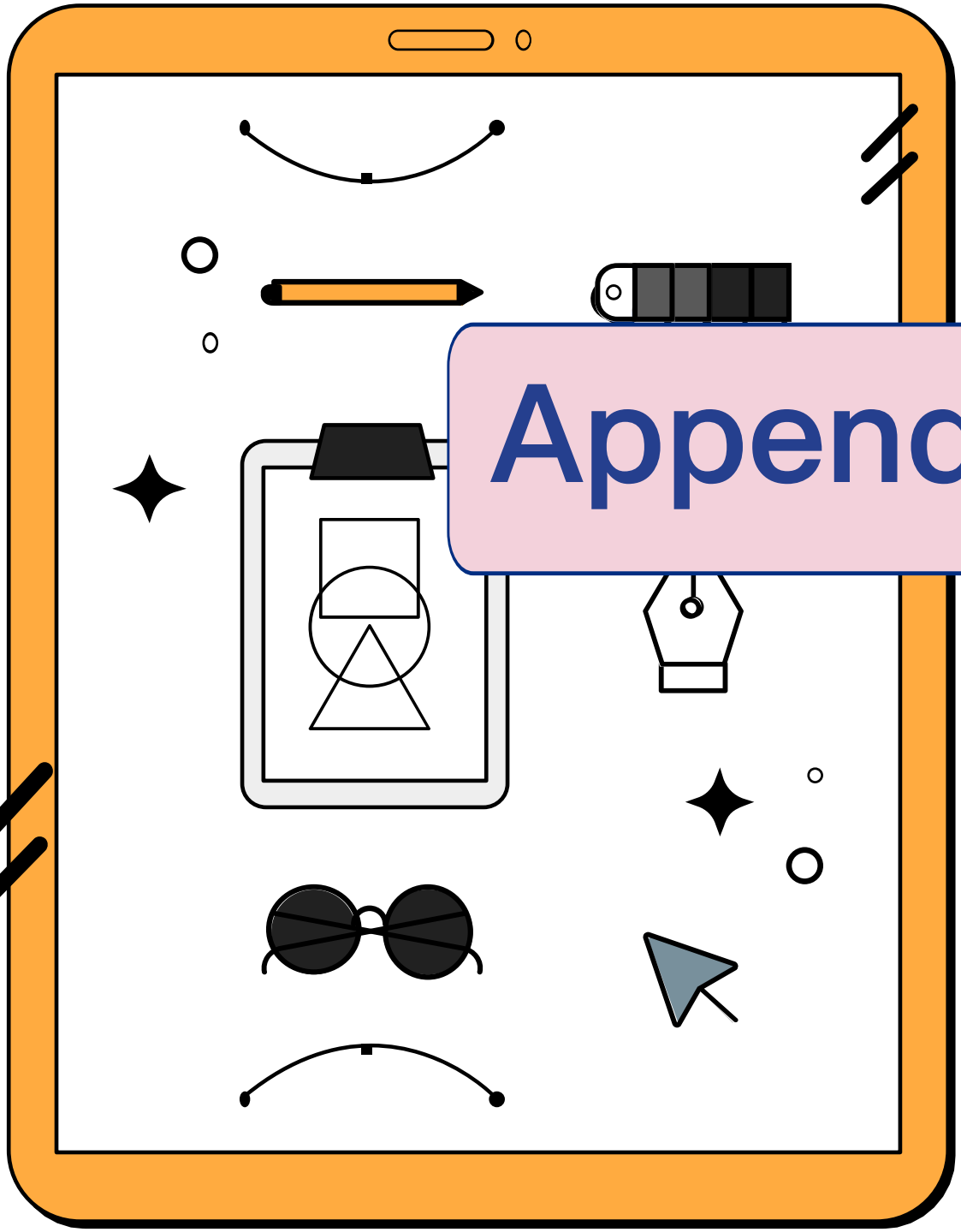
The more reasonable the context is, the better designers could work.

Ineffective processes for the whole flow

UI related to use case

What is the scope of the prototype?

1.1.7 Also as you know, I'm an IPD major and I'm not very familiar with creating this kind of interactions. ai products are also more than just internet products, there will be some physical interactions as well. Can I use it to create product physical interactions in the future? I don't think so far it seems to be able to.



Appendix H

Test plan & Materials for Test 2

Introduction

"Welcome to the user test! The project is crafting a digital tool for designers to create interactive behaviors between humans and AI in specific user-tailored situations.

The primary aim of this test is to assess whether the current prototype facilitates designers in comprehending specific terms and enables them to easily complete straightforward design tasks.

This tool is usually used to create conversations between humans and AI by sparking interesting thoughts. Unlike other tools, it doesn't make things like buttons for users to click on.

There would be 3 parts. The first part is one short interview about your basic information. The second part has 3 tasks, you are invited to compete them and answer some questions. The final part is a short interview about your experience on using the prototype.

During the process, it would be the best if you think it out. And I would record screens and audios.

Because it is one pilot test, there might be something running not logically. If you have any suggestion on the test or on the prototype, be free to speak it out during the test or one interview after the test.

Do you have any question till now?"

For Jiayi

Pre-task-interview

1. "How long do you study for design?"
2. "What kinds of tools do you always use for the design process, like persona, user journey map, empathy map and so on?"
3. "Do you have experience in design products related to AI?"
4. "What kind of digital tools or products have you used to help you design interaction between humans and AI?"

After-task-interview

1. Explain why you choose to add the message in that way.
2. What do you think of the experience when you create a new blank message, fill in the message and then make interactions? Does it work as you expect?
3. Is there anything you think the system could develop? For example, is there anything would be better?
4. Is there anything else you want to share?

Additional during each task

If Participants miss some important functions, guide participants to click them.

For Jiayi

Test plan

A test plan is created to prepare for user testing. The test is intended to see if the model design meets the design goals and to identify areas where we can still make improvements. The test plan is shown in the figure.

Procedure
There are a total of 6 planned steps for user testing. The entire test process is expected to last 1 hour.

Participants
To get the best results, it would be ideal to test the prototype with 5-15 participants.

Equipment and setup
Record tool: The laptop for participants; QuickTime Player.
Test tool: Laptop with Figma

Responsibilities
The facilitator will make introductions and begin the interview. As the participants begin to complete the task, the moderator will turn on the QuickTime Player to record the participants' behavior and voice. At the same time, the facilitator will play the role of a Figma prototype monitor to make sure that the prototype goes well.

For Jiayi

Tasks

Imagine you're like a superhero **designer,** and you have a special tool on your computer to help you create amazing things. Your mission is to use this tool to solve a puzzle about how people and AI communicate to each other based on a specific use case. Can you do it? Ready, set, design! 🚀

Some terms that may help you explore accomplishing this mission: See Term cards.

Use case: See the poster & the storyboard 1

Task 1: Here are 2 examples of ***Message*** on one digital prototype based on the use case and the storyboard 1 for the HR manager.

Now you can spend ***5 minutes*** trying to answer a few following questions by exploring this interface:

- What do these two examples describe? Type your answers on the prototype on each message.
- Tell me out loud what the respective roles of the human/AI in the message were and what kind of information exchange took place?

There are some explanations on some parts in the prototype, like "Creating new roles", "Viewing details of types of information" and so on . If you are interested, you could have a try.

- After you explore on this interface, what else do you want to share with us?

Task 2: Here is one example of one ***sequence of Messages*** on one digital prototype based on the use case and the storyboard 1 for the Hr manager.

Now you can spend ***10-15 minutes*** trying to answer a few following questions by exploring this interface:

- What does the sequence describe? Please tell me out loud what role the human or AI plays in each message and what information is exchanged.
- Are you able to find the corresponding story scene in Storyboard 1 for each message?
- What do you think of the "Interaction Row" while you are trying to understand the specific sequence? Does it help you understand the prototype or the message sequence?

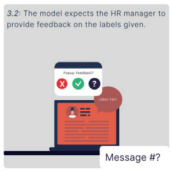
There are some interesting functions like "Adding a new blank message", "Changing message orders", "Viewing details on Interaction Row" and so on. If you are interested, you could have a try.

- After you explore on this interface, what else do you want to share with us?

Task 3: Below is an example information sequence on a digital prototype showing what is in the storyboard 2. Now, the design team believes that adding step 3.2 will help improve the overall user experience.

Now, create the ***Message*** that belongs in the digital prototype according to step 3.2 in Storyboard 2.

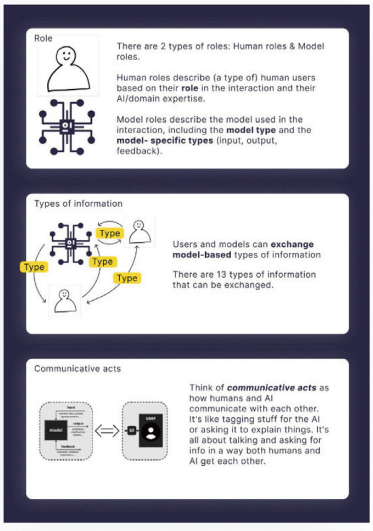
Due to prototype development limitations, the prototype will only display the corresponding content when you click on the preset button.

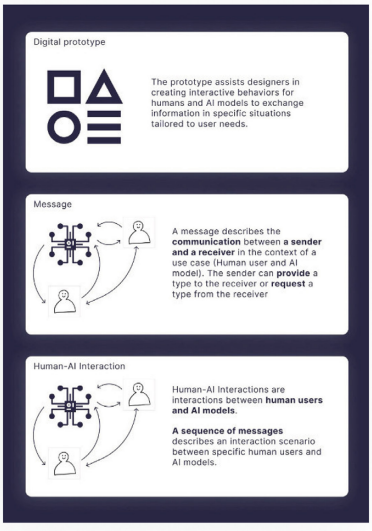


Step 3.2

There are some helpful cues:

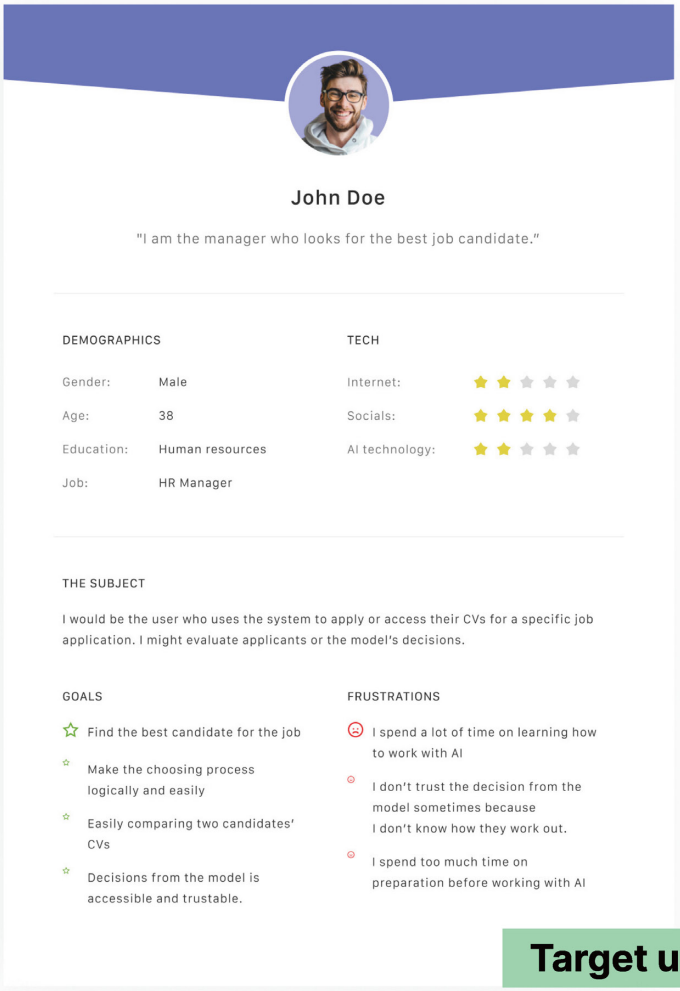
- Create a new blank message
- Fill the message based on the Step 3.2 and the Description
- ...



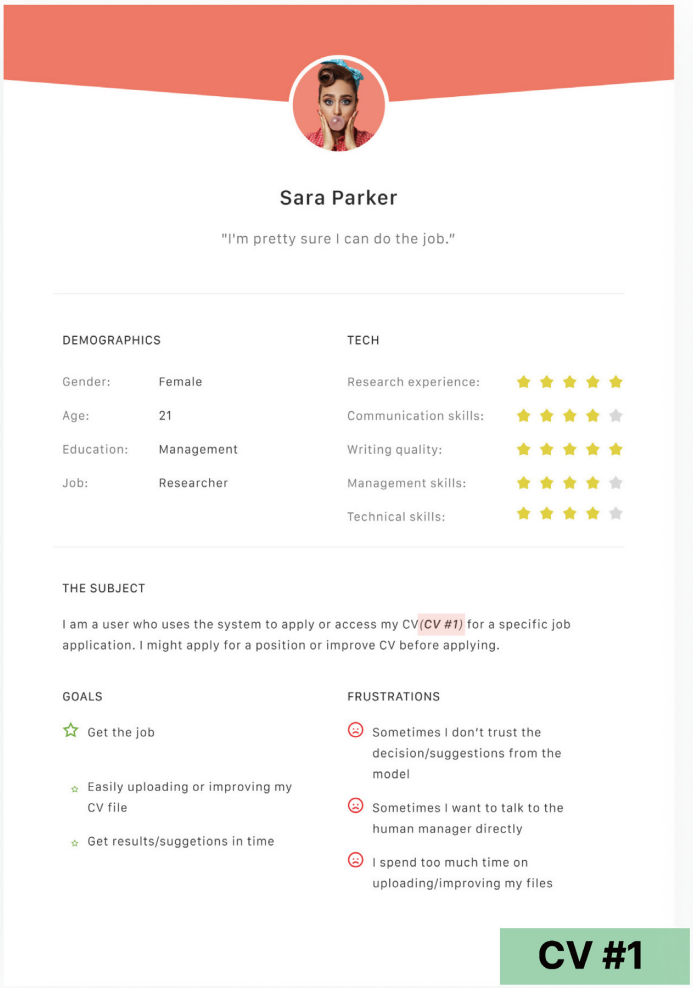


Materials for understanding the use case contest

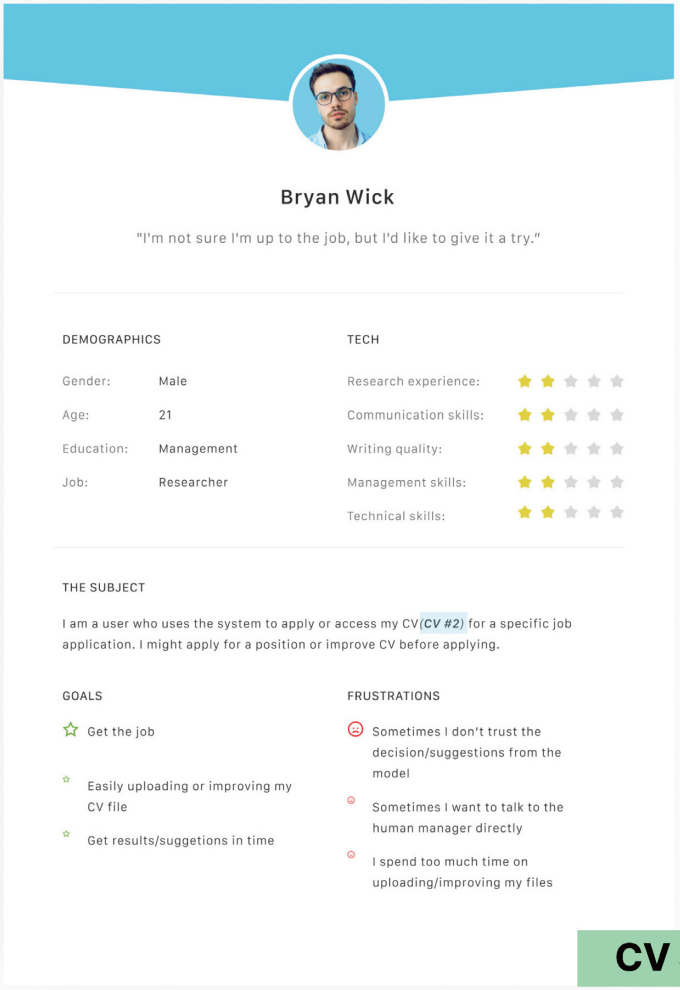
Poster



Target user



CV #1



CV #2

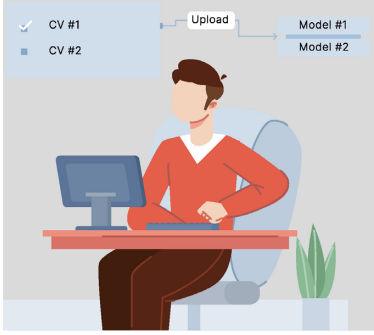
Storyboards for Task 2 & 3

Storyboard 1

View this storyboard illustrating how an HR Manager engages with two AIs to address a dilemma when faced with difficulty in selecting a candidate.

1. The HR manager wants to know who would be a better candidate to get the job.

2. The HR manager uploads CV #1 into the AI systems to get suggestions.



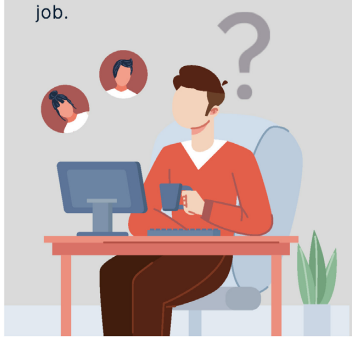
3. Models work and provide suggestions in the form of positive or negative labels for the CV #1.

4. The HR Manager is satisfied with the labels afterwards.

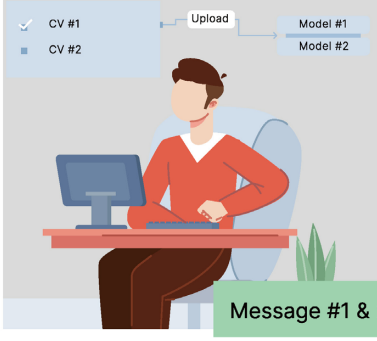
Storyboard 2

The storyboard has the same content as the previous one. But each step corresponds to the serial number of the message in the prototype. hopefully helping you understand the relationship between the message and the storyboard.

1. The HR manager wants to know who would be a better candidate to get the job.



2. The HR manager uploads CV #1 into the AI systems to get suggestions.



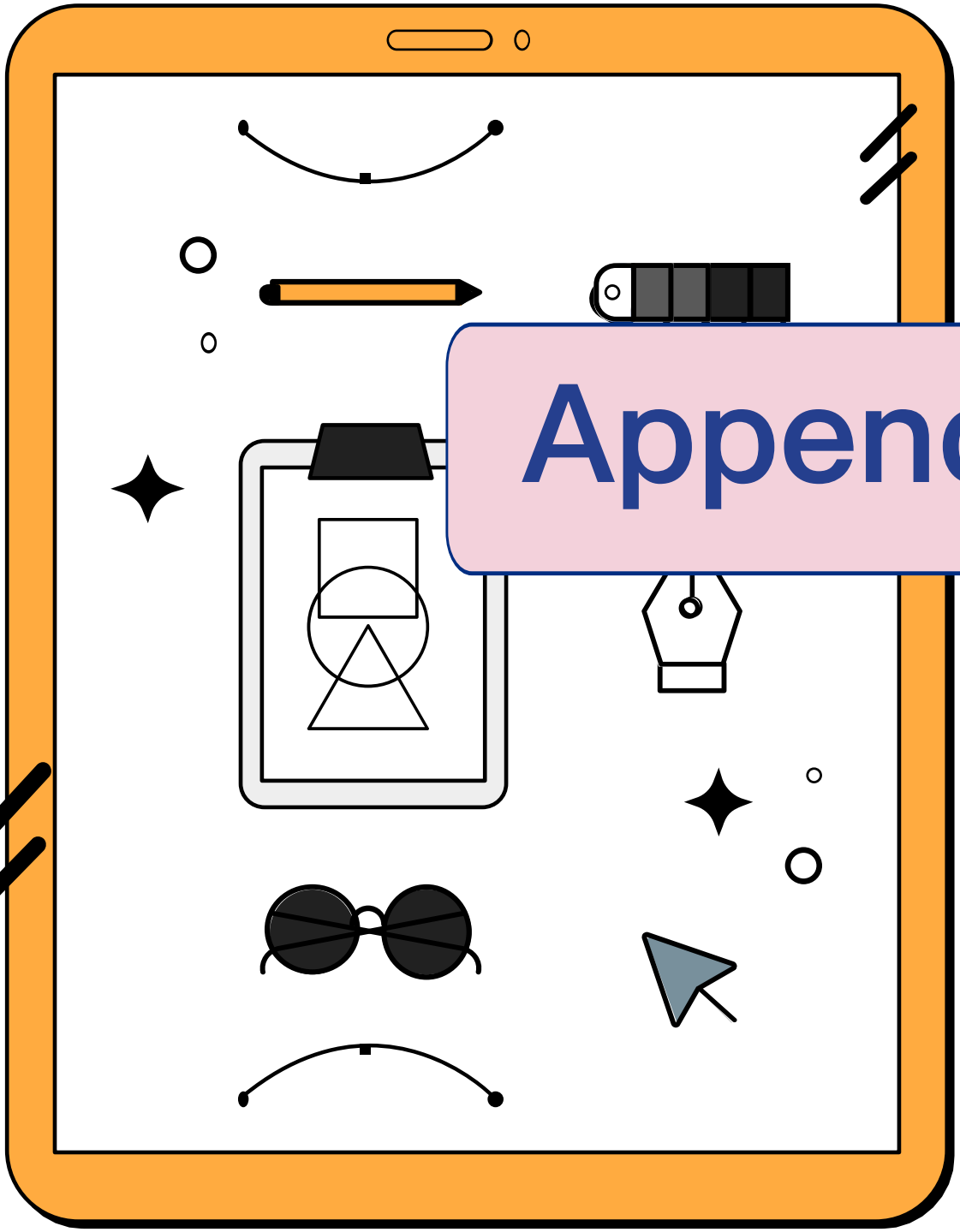
Message #1 & 2

3. Models work and provide suggestions in the form of positive or negative labels for the CV #1.

Message #3

4. The HR Manager is satisfied with the labels afterwards.

Message #4



Appendix I

Transcription for Participant 1

Pre-task-interview

Jiayi Zhou Participants

1. "How long do you study for design?"
More than 5 years. And now I am studying as one DFI student.

2. "What kinds of tools do you always use for the design process, like persona, user journey map, empathy map and so on?"
Yes, I have rich experience in using all these tools, but to be honest I don't trust some of them. I love the user journey map because it could work as one visual picture to summarize milestones. It helps me remember what my target groups might do and how they experience. However, for tools like personas, I don't like it a lot.

3. "Do you have experience in design products related to AI?"
Yes, the ITD class. My group collaborated with an automotive brand to design a car interior for human-AI interaction through sound.

4. "What kind of digital tools or products have you used to help you design interaction between humans and AI?"
For design, of course Figma. But for prototyping, I would say Protopie.

Fixed Provide/ Request button

Assessing actions with Active-Passive Relationships

Unclear on the mental or the physical behaviours in storyboard

Examples make it confusing.

Didn't see the type of information now.

The two sections look like they were chosen for different audiences

Task 1

Jiayi Zhou Participants

When I first saw the interface, I was struck by "Provide" tab. What is "Switch to Request"? (Figure 1.1)This piqued my curiosity, so I clicked on it to see what would happen. (Silence) "Do you have anything to say about this place? It looks like you're very hesitant." I find this place a bit strange. Clicking here seems to change the text in the button directly, and I'm guessing that affects some of the functionality as well. See Figure 1.1. Because my task is to go and determine what this (first) message example conveys the meaning and find the step on the Storyboard that corresponds to it. "Provide" and "Request" can to some extent be interpreted as two behaviors that may correspond to different steps on the storyboard. "So you think not being able to change the state of "Provide" or "Request" would help you better understand what each message is trying to convey, right?" Yes. "Request" means that the manager wants the model to give him something or information, while "Provide" means that the manager has given the model some information on his own initiative. So the active and passive relationships would be completely reversed.

(Check the storyboard #1) → Can I ask some questions about Storyboard content? For example, in the last step, does this manager need to do any physical behavior to communicate with the machine? Or is this all about the manager's mental work? "The HR Manager needs to exchange information with the model." OK. What about the first step? "This is a pure Mental Work. This step shows the problem faced by the HR Manager. This step shows the problem faced by the HR Manager. The second and all subsequent steps are based on finding the answer to this problem, which results in the interaction and exchange of information between the HR Manager and the AI." OK. Based on the information I've gotten so far, I think the first message is probably from Step2 or 4 (Figure 1.2). Because this message says "Provide", but it doesn't say what the Provide thing is. The step in the storyboard that involves the manager proactively giving the model should be either 2 or 4. For the second message, I also think it comes from either step 2 or 4, favoring 2 (Figure 1.3).

"Did you notice that there is a "New sample" underneath? Do you feel that this section is integral to what we just discussed?" Before you told me just now, I didn't notice it at all. What is the meaning of 'Create a new feature vector'? See Figure 1.4. "You could click it to get the answer to the question." (After reading the explanation) So you're wondering if I can look at a column of information as a whole, right?" "Yes, from your perspective as a designer, can you habitually think of this column as a whole? Or do you think is row of content is more likely to be perceived as building the same message?" In terms of this information being designed as a table, I can very easily understand that the contents of a column are forming the same information as a user journey map. But what I find confusing is that both "Provide" and "Request" are describing the behavior of the manager. In the "Type of information" line, all the information gives me the feeling of an information description developed specifically for the machine, as if describing how the model handles the information exchanged between the human and it. So overall, my understanding of the second line is that it describes "this is describing how a person communicates information to the AI", while the third line describes "how the machine processes the information exchanged by the person".

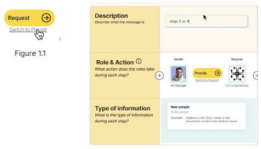


Figure 1.1

Figure 1.2

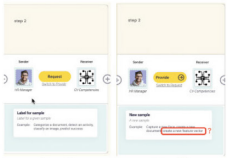


Figure 1.3

Figure 1.4

Confused on the function limitation of the prototype

Confused on how the model role/actor role could work

Don't understand the function of each part in the prototype.

Need clear explanation on what the prototype could do and more things on roles.

Needs for viewing the prototype playing

(Time for freely exploring) Exploring creating new roles → I don't think there's anything here that particularly confuses me. But I do have a question. I'm not sure if the models in these (Figure 1.5) need to be strung together by me and then have them go attached to my technical needs. Is this achievable? In other words, do I as a user need to select these modules? Because I haven't researched the behavior of hiring managers, I don't know their work patterns very well. If I were them, I would need the first two modeling features. And I don't want to know information about the process in the middle of the model, I just need it to give me the final answer. See Figure 1.6. I also have another question. Do you know what will happen if I create a new role? For example, if I create a new character, what does that new character mean to the mod? I might ponder this question. And there would be another problem. If I create a few model role, If I create some models, does their ordering in the right toolbar affect the functionality? For example, is the working order top to bottom? "The idea we have now is that the right side just displays the character library, and you can fill it in by selecting the desired character from the library as you fill in the blank information. At the same time this mod doesn't work in the prototype. It's still textual information that visualizes the overall interaction logic." Ok. "So does it make sense now?" So far it's reasonable. But I'm curious as to how the form of interaction I'm creating will be presented? Is there a button that when I click on it, it will run the overall message or will he only run specific messages? How will it render the interaction? "Currently the prototype does not allow designers to visualize how their model will work, but there will be a preliminary idea in a follow-up task that may help designers see the results." Okay. Other than that, I think everything else is in line with what I'm thinking so far. I'm very curious about how the different AIs are going to be cascaded, and whether they work independently or can work together.

Figure 1.5

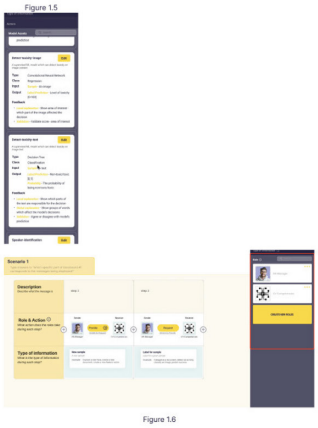


Figure 1.6

The confusion on what the prototype could do.

More direct and easy-to-understand content

Confused on what they could do.

Input/Output/UI elements belong to different architectures.

Task 2

Jiayi Zhou Participants

When I see the table again, I will read it from top to bottom. Since the right side shows that there's more of this table obscured, I guess I can drag this panel left and right to move it around. But as I thought about the process, there was one place where it wasn't quite what I had intuited, or expected, it would be. Because when I look at a watch, I look at it from top to bottom and left to right. So the first step is to upload CVs, and the second step is to go against that CV and set the label I want. but I'm curious what the process looks like if it's an application? "Can you explain the problem more?" For example, whether there are uploaded files or not, there will be a list of features here that I can then manipulate. Or I need to upload a file first, and the model generates me features based on that file that I can manipulate, and then I can select them. "You can click on the mod's details screen and there will be some information here that might answer your question." See Figure 1.7. Make sense.

The third message is very obvious. The fourth one seems a bit odd to me, and there seems to be a problem with the prototype. Based on the first quest and the information brought to me earlier in this quest, the Description and Actor&Actions content in the fourth doesn't match. See Figure 1.8. If based on Description, I'd file this info under step 3 in the storyboard. But based on the follow-up information, I think it belongs in Step Four. "Where do you think the exchange of information begins and ends in this sequence?" It obviously starts at step 2 of the storyboard and ends at step 4. Meanwhile the table goes from left to right and exchanges information.

"There are other features of this interface that you want to keep exploring? It might help with the third quest." I first wanted to just click the interaction button. This looks like the gateway to a whole new world. See Figure 1.9. (After Jiayi showed how to create a new blank message and how to change message orders) I didn't realize I could do these things. But it seemed very reasonable. See Figure 1.10.

"Now let's start exploring the Interaction Interface. Remember in Task 1 when you asked how to PREVIEW the human-AI interaction created by the designer? Well, you may have your answer here, as the Interaction Row section will present input/output/UI elements to help you understand the information in the use case and how the vehicles involved play a role in the exchange of information." Well, I see. It is very clear. What makes me confused is still the same as Figure 1.8. "And do you think this view (which has helped you understand how information is exchanged)?" In doing this task, I felt very clear. But if I were to do other use cases independently, I would be confused by the words "Input/Output/UI elements" See Figure 1.11. because they give me the impression that they don't belong to the same category. But since I already have a better understanding of the above information due to the task I just did, I don't focus too much on these questions, but on what's inside of them, so it's very clear to me. So the current content of the different information layers within this line makes me feel that it may be related to the needs of the target group, possibly at the direction of the designers based on the needs of the users. "How you could click somewhere to find explanations on Interaction Row." See Figure 1.12. It is easy for me to find explanations, but I think these terms should be intuitive and easy for designers to understand before they look for more detailed explanations. How do you put all these seemingly separate things in one line? Why do you do this? It feels like it needs to be presented in an easy to understand way.

Transcription for Participant 1

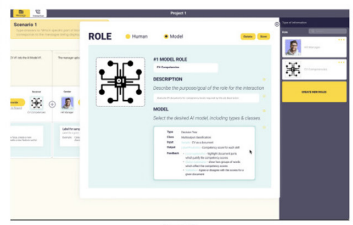


Figure 1.7

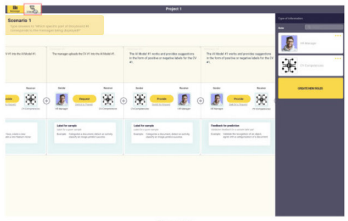


Figure 1.9

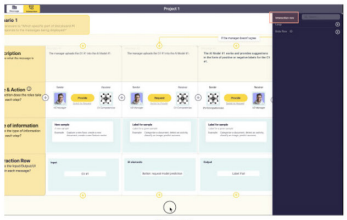


Figure 1.11

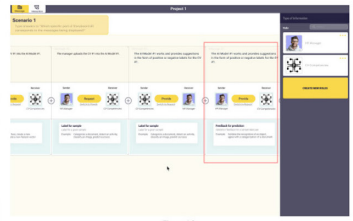


Figure 1.8

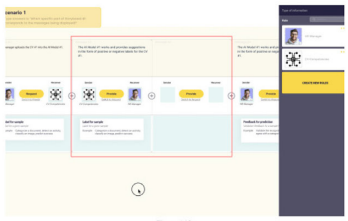


Figure 1.10

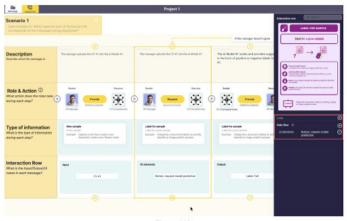


Figure 1.12

Different icons for different functions.

What does the yellow plus button mean? See Figure 1.13. "When designers are working on a task, a linear process will help them accomplish it more efficiently. However, loops may form at some small pivot points in linear tasks. So the yellow plus button is a way to give designers a way to connect two pieces of information that are not adjacent to each other, thus forming a loop. It can also be understood as a case of icon-format judgment, if statements, and so on."
Oke. My suggestion would be to use another icon or presentation to differentiate between adding new blank information and adding a connection.

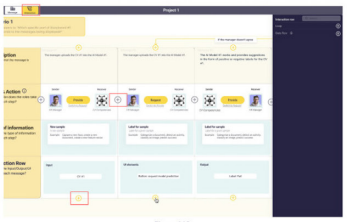


Figure 1.13

Change the UI to hide the explanation

The same way to interact in the same goal

The place of Type of Information is not clear.

Designers have different ideas on defining Provide & Request.

Pay more attention to details

Task 3

Jiayi Zhou Participants

(After creating a new message) What should I do to now? How can I fill in the message? (Click the sender blue square) There is nothing happening. May I click the right function part? Ok, it works. Now I need to choose actions. See Figure 1.14. Maybe request... No I think it should be the Provide. Honestly, I think both behaviors are fine, it's all about how the designer interprets the behavior. Request puts more emphasis on the AI actively wanting HR to do something, but for me, I don't think the AI is in a position to make people actively do anything. It should just be an assistant to help HR to deal with the demand.

Now what should I do to fill in the information part? (Click the blue square and view the list of information) I think it should be the label for sample. Nothing happen. I don't do the right thing. "Why do you think it is the Label for sample?" Maybe I need the explanation of each information. Where could I get that? The yellow one? Ok, yes. (After reading details) "Why do you think it is the functions for providing?" I'm a little confused on the type of information step, should I put in the information that the model wants to get or the information that the model should deliver? "From the experience of the previous tasks, what kind of interface do you think is reasonable?" For me, I only want the HR to agree or disagree with labels. The type of information is in more abstract levels. I pay more attention to details. Is there the information about Agree or Disagree? "For Type of information, there are choices including Agree or Disagree but including other design insights at the same time. If you want to find one type of information including only Agree or Disagree, no. Do you think it is related to something have like in the Interaction Row?" It is a little abstract for me. I think it is the best if we could have space to view Agree or Disagree. Like the version in Task 2. Yes, after clicking the Interaction part, I see it. But I hope to visualize it more in detail. For example, I could give HR two buttons to represent Agree or not. Or I could ask HR to report his ideas in another interface. You could see in Figure 1.15, these are the same thing for me. At least they will be presented in the same interface, if I am the designer. I didn't have any idea if there would be something like in Figure 1.16.

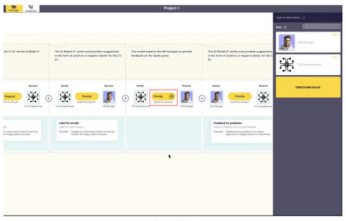


Figure 1.14

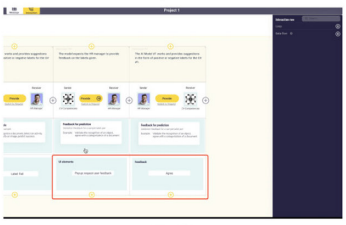


Figure 1.15

Transcription for Participant 1

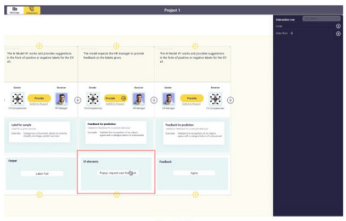


Figure 1.16

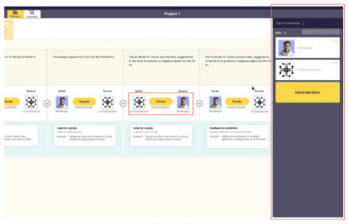


Figure 1.17.1

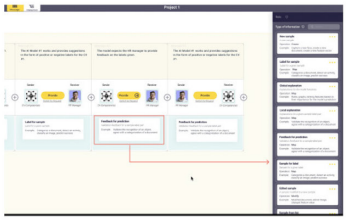


Figure 1.17.2

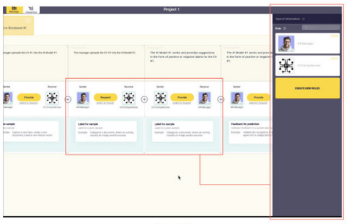


Figure 1.18

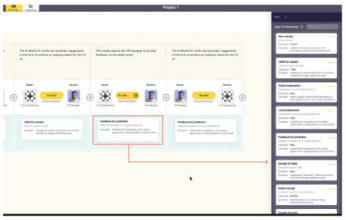


Figure 1.19

Transcription for Participant 2

Pre-task-interview

Jiayi Zhou Participants

1. "How long do you study for design?"
More than 5 years. And now I am studying as one DFI student.

2. "What kinds of tools do you always use for the design process, like persona, user journey map, empathy map and so on?"
Yes, I have rich experience in using all these tools, especially the persona and the user journey map. Almost all my projects would use the both tools.

3. "Do you have experience in design products related to AI?"
Yes, a lot. For example, the ITD class and one elective class.

4. "What kind of digital tools or products have you used to help you design interaction between humans and AI?"
In the past, Figma. And Voiceflow.

Task 1

Jiayi Zhou Participants

Let me see. The first row is about Description, I think that is the place for me to type my answer. The second row shows two different roles. Ok, it is the place for me to explore what the message is, right? And how can I interact with the prototype? I find I cannot click the Switch or the Provide part to change the button. Is that correct? "Yes. Because that part works for the answer to Question 1, it keeps you easily to understand the example of what could change the button, the description of the message could change." Make sense. There are only two roles "Sender" and "Receiver" here, so is there another role? "Only two." Ok. Let's talk about the first question. To be honest, I am not sure if I find the message. You tell me there are two examples of messages and here are truly something related to two messages. I am sure the right part is a zone for stakeholders. Based on the term you introduced for me, I think the type of information is the message (Figure 2.1). It is the information exchanging during the communication between HR and Model in this step. What does it 'create a new feature vector' mean? Is it something the model would do during the interaction or after getting the information? See Figure 2.2. "It is the example to explain what new sample is." It is confusing for me.

Let me write down what happen. See Figure 2.3. Wait! For the second part in Figure 2.3, since you set the Request as a message that can't be altered, I again think that perhaps the message here describes HR asking the model to classify and give results. Now I prefer the new description.

"Do you remember before the test, I introduce the message? A message specifies the sender or receiver and the type of information involved. So now, how do you think of the examples?" Yes, now everything is clear for me.I almost forget the definitions although you introduced in the beginning.. Another reason why I wasn't sure about MESSAGE before was the second question. The second question as I understand it is for me to decide by myself. I thought the task was subjective.

Another thing that may have led me to only think of Type of information as a complete message in the first place is the fact that Chatgpt's interactions are now based on conversations. So this model has led me to already be somewhat prejudiced against the way people interact with AI, i.e. only through text. "So do you consider Type of information to be information created based on machine understanding?" It depends. For the examples, I think they are from humans. It is also reasonable if it is from AIs, like informing humans.

"Here are some explanations on types of information, if you want to know, you could have a try." Ok, now it makes the example more reasonable. For the Figure 2.4, I think the words could be developed. In New example part, the two places are almost the same.

See Figure 2.5. Is it a button? For me, it is not like a button, because there is one icon nearby. And it's not in a location that's easy to notice. A lot of them I couldn't understand what they meant by just reading the brief introductions. How do I go about getting more information? This yellow icon? It feels a bit inconspicuous. Couldn't these be arranged in some sort of order? Right now they seem to have different colors, but not sure what that means.

Fixed Provide/ Request button

Hard to understand what the message is even we have introduced it

Examples make concepts hard to understand

Fixed Provide/ Request help participants understand it

The second question might be confusing.

Make the prototype's goal clear and keep them far way from the current tools

More visualization elements on UI

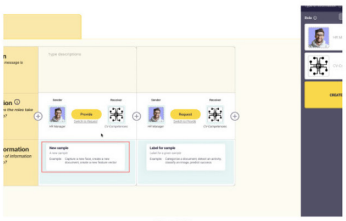


Figure 2.1

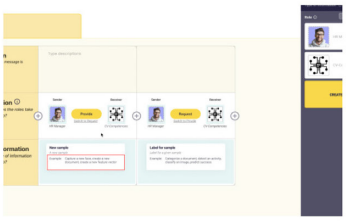


Figure 2.2

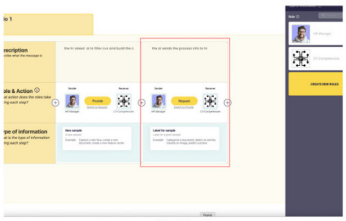


Figure 2.3

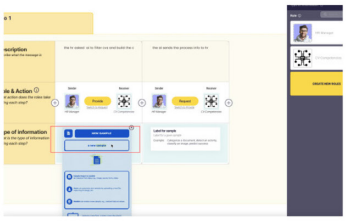


Figure 2.4

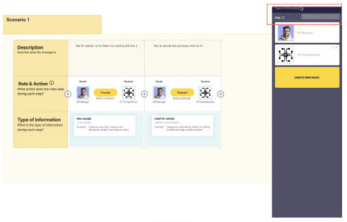


Figure 2.5

Figure 2.6

"Is there anything more you want to share?" Yes, the message now is not introduced well and clear. Perhaps you could explain it better with a newbie tutorial. Secondly, I think there are some things that could be done better with the presentation of information here as well. For example, showing the main body of the message at the beginning (Figure 2.7), rather than the top line being just for me to fill in the content. I think it would make more sense to move the Description line to the bottom.

Thirdly, see Figure 2.6. Now the interpretation of information does not exist at the same time. For example, I would like to see both the New sample and Sample for Label explanations to help me make a choice, which I can't do now.

Other way to introduce concepts and functions

Show important information first

Bigger spaces for Type of information & show them at the same time

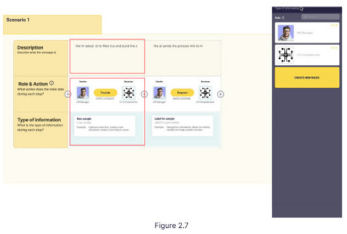


Figure 2.7

Transcription for Participant 2

Task 2

Jiayi Zhou Participants

After the Task 1, now everything is clear for me. I could click the Provide, right? Oh, it switches to Request! See Figure 2.8. Why is there one icon? Oh, it works as Confirm!

Then I want to click the Interaction button. Wow what is the interaction? Ok, I get it! It is something like the Touch point in the user journey map, right? See Figure 2.9. All of these things are points that the designer will address in the interaction of exchanging information for Human as well as AI, such as the items being exchanged or the triggers that trigger the exchange behavior.

"Does the Interaction Row help you understand the sequence or how the prototype works?"
No. To be honest, I have ideas on how the sequence goes and how the prototype works, so it doesn't help me on that. However, it helped me visualize this process of information exchange more concretely.

It is really like a user journey map. Or at least it reminded me of this user journey map. and I was able to understand the prototype somewhat quickly by learning and applying the user journey map before.

Is there any button to add a new blank message? Ok. Maybe it would be easier to notice it if you put it in the first or second plus button.

What's the difference between Message and Interaction button? Oh I see, there would be more details in Interaction Interface.

But I still don't like the placement of Description. People are visual creatures, and are more concerned about the information conveyed visually.

What will happen if I click the yellow plus button? Oh, I could connect messages in other places like IF in coding. Like one Mind-map!

One more suggestion I want to share is right now this interface feels a bit limited. Wouldn't the tool work if the design process evolved according to the "two-diamond model"? Also I'm more interested in learning more about the 12 types of information just mentioned in the Interaction interface. But I find there is nothing.

After this task, I think I understand what the prototype is. It might work for the brainstorming part. And based on this recognization, I think the top important information should be the 12 types of information. I also have a great need to know this when I fill out a message or do brainstorming.

Good word. Touch point is a good word to present what input/output/UI elements and it is easy for designers to understand

The function is to visualize the instances which trigger messages flow from the left to the right

An understanding of the user map can go some way to helping designers understand the overall interactive tool

Put the clickable button on the first/second button

Suggestions on information structure on message content

Understand when to use the prototype and how the prototype would work

Don't get why there are links and yellow plus buttons

Pay more attention to Type of information

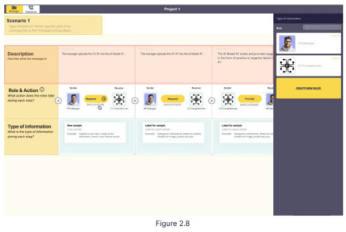


Figure 2.8

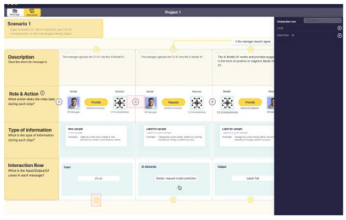


Figure 2.9

Task 3

Jiayi Zhou Participants

It is to add one feedback. To be honest, I think there should be one new feedback after each message. Oh, I need to follow the task guidance to add the specific feedback.

(Add a new blank message) What should I do now? Fill in the sender? (Click the sender blue square) Confusing. Oh, I need to click the roles from the right part. Based on the description, I think the sender is the model.

(Forget to change the Provide) "So you think it is the Provide here?" I think it's Provide because it's the model coming forward and offering it to HR. How is HR going to provide feedback if the model doesn't provide HR with a popup or something? See Figure 2.10.

"I want to ask more about your opinion on Request and Provide. How do you understand them?" Passive is Provide, active is Request. For me, I've always seen people in a passive position when it comes to communicating with AI. Even if the person provides the model with any information in the beginning, it is because the model provides a certain information interface. Simply put, a machine thinks a human can give it before a human can give it.

"So how do you define Request or Provide in one description? " Good question. See Figure 2.11. I think the start is always the model request the HR to give some information. And the Message #2 should change the Request into the Provide. After that the machine could request others.

Designers could have different definition on provide or request.

Different designers have different ideas on Provide & Request.

Define clearly what Provide & Request is.

Figure 2.10

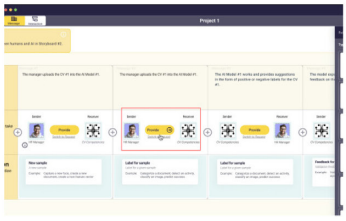


Figure 2.11

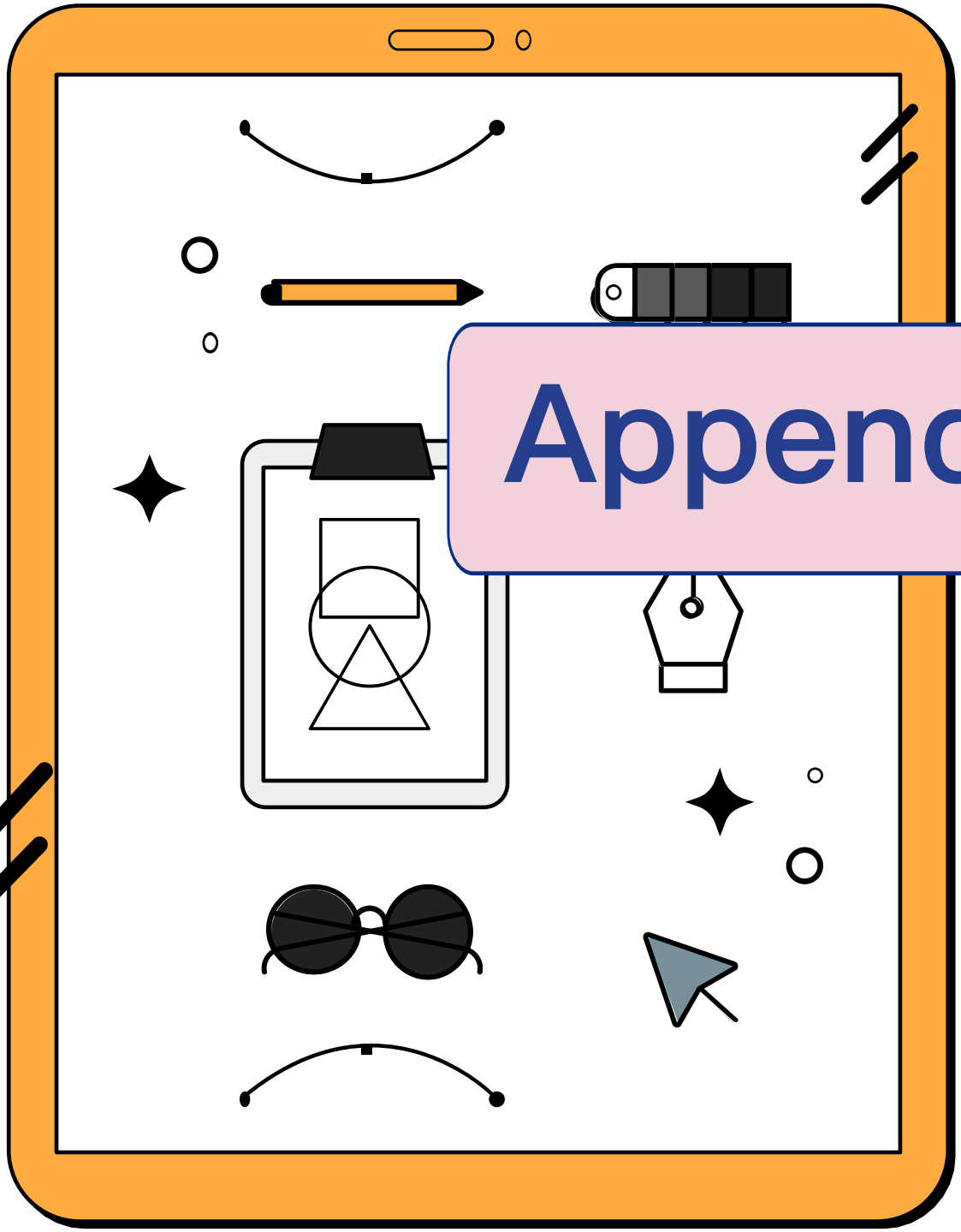
After-Interview

Jiayi Zhou Participants

1. What do you think of the experience when you create a new blank message, fill in the message and then make interactions? Does it work as you expect?
It works smoothly for me until now.

2. Is there anything you think the system could develop? For example, is there anything would be better?
Until now, no.

3. "Is there anything else you want to share?"
No.



Appendix J

Test plan & Materials for Final test

Introduction

"Welcome to the user test! The project is crafting a digital tool for designers to create interactive behaviors between humans and AI in specific user-tailored situations.

The primary aim of this test is to assess whether the current prototype facilitates designers in comprehending specific terms and enables them to easily complete straightforward design tasks.

This tool is usually used to create conversations between humans and AI by sparking interesting thoughts. Unlike other tools, it doesn't make things like buttons for users to click on.

There would be 3 parts. The first part is one short interview about your basic information. The second part has 3 tasks, you are invited to compete them and answer some questions. The final part is a short interview about your experience on using the prototype.

During the process, it would be the best if you think it out. And I would record screens and audios.

Because it is one pilot test, there might be something running not logically. If you have any suggestion on the test or on the prototype, be free to speak it out during the test or one interview after the test.

Do you have any question till now?"

For Jiayi

Pre-task-interview

1. "How long do you study for design?"
2. "What kinds of tools do you always use for the design process, like persona, user journey map, empathy map and so on?"
3. "Do you have experience in design products related to AI?"
4. "What kind of digital tools or products have you used to help you design interaction between humans and AI?"

After-task-interview

1. How do you think of the Touchpoint?
2. What do you think of the experience when you create a new blank message, fill in the message and then make Touchpoint? Does it work as you expect?
3. Is there anything you think the system could develop? For example, is there anything would be better?
4. Is there anything else you want to share?

For Jiayi

Test plan

A test plan is created to prepare for user testing. The test is intended to see if the model design meets the design goals and to identify areas where we can still make improvements. The test plan is shown in the figure.

Procedure
There are a total of 6 planned steps for user testing. The entire test process is expected to last 54 - 90 minutes.

Participants
To get the best results, it would be ideal to test the prototype with 6 participants.

Equipment and setup
Record tool: The laptop for participants; QuickTime Player.
Test tool: Laptop with Figma

Responsibilities
The facilitator will make introductions and begin the interview. As the participants begin to complete the task, the moderator will turn on the QuickTime Player to record the participants' behavior and voice. At the same time, the facilitator will play the role of a Figma prototype monitor to make sure that the prototype goes well.

For Jiayi

Tasks

Imagine you're like a superhero **designer,** and you have a special tool on your computer to help you create amazing things. Your mission is to use this tool to solve a puzzle about how people and AI communicate to each other based on a specific use case. Can you do it? Ready, set, design! 🚀

Some terms that may help you explore accomplishing this mission: See Term cards.

Use case: See the poster & the storyboard 1

Task 1: Here are 2 examples of ***Message*** on one digital prototype based on the use case and the storyboard 1 for the HR manager.

To make sure that each message expresses only one kind of content related to storyboard 1, some buttons like "Provide/ Request" have been fixed.

Now you can spend 20-25 ***minutes*** trying to answer a few following questions by exploring this interface:

- Are you able to realize that those content of the digital model you are looking at constitute Message #1? What are the content of Message #2?
- What do these two examples describe? Type your answers on the prototype on each message and tell the researcher the corresponding story scene in Storyboard 1 for each message.
- What's the meaning of one type of information called "Prediction probability"?
- If you want to change the function module of Model 1 to "CV-Screening", how should you do it?
- After you change Model 1's function, are there any other changes either?

Task 2: Below is an example of an information sequence for a digital prototype that illustrates the HR Manager's use case and a portion of Storyboard 1.

Now you can spend ***20-25 minutes*** trying to answer a few following questions by exploring this interface:

- Message #3 & 4 seem to be in a wrong order. Please switch their orders.
- What does the sequence describe? Please tell me out loud what role the human or AI plays in each message and what information is exchanged.
- Are you able to find the corresponding story scene in Storyboard 1 for each message?
- Do you know how to create or delete a new blank message between Message #1 and Message #2?

Go to the Interaction interface. You can notice that the Touchpoint in each message doesn't seem to match the description in the Description. Please combine the storyboard content with the description in each Description and modify each Touchpoint to correspond.

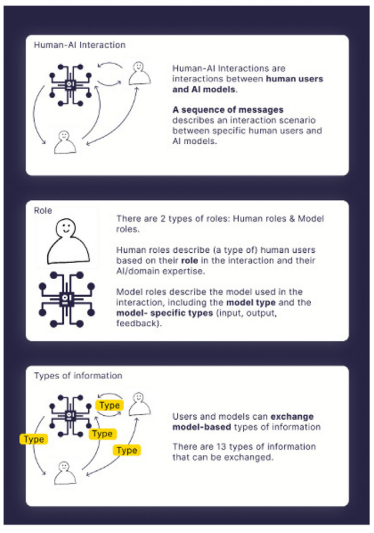
Task 3: Now, building on the Message sequence of Task 2, the design team felt that adding Step 3.2 would help improve the overall user experience. See storyboard 2.

Now, create the "Message" that belongs to the digital prototype and the "Touchpoint" that belongs to it according to step 3.2 in "Storyboard 2".

Due to prototype development limitations, the prototype will only display the corresponding content when you click on the preset button.

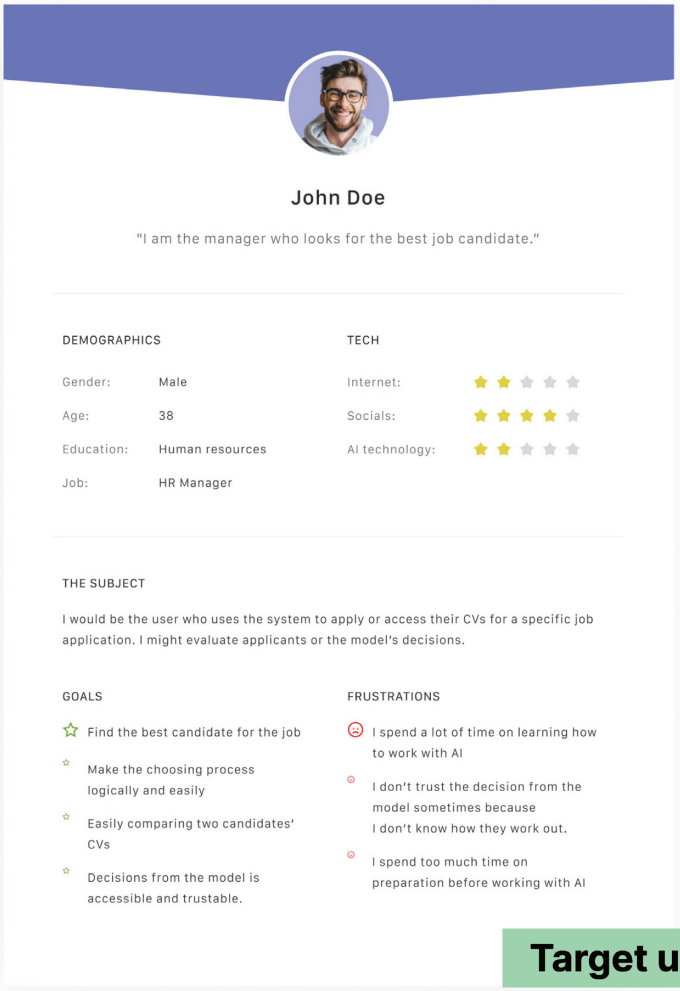
There are some helpful cues:

- Create and fill a new message based on Step 5.
- The Touchpoint in Step 5 is a Popup, and the content of the Popup is to ask the HR Manager if he/she agrees with the prophecy.

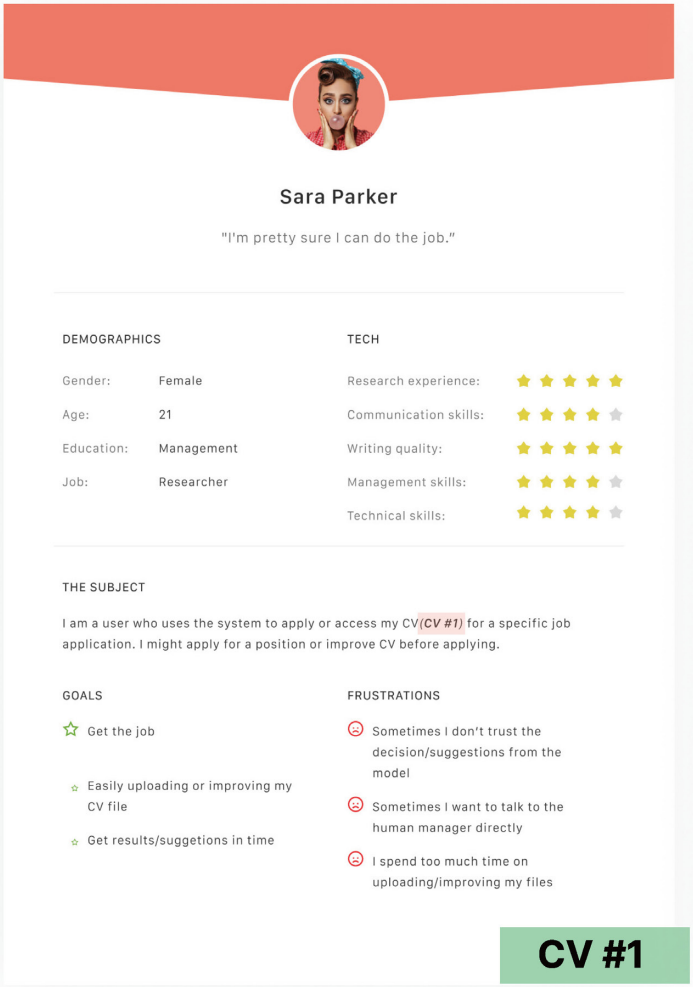


Materials for understanding the use case contest

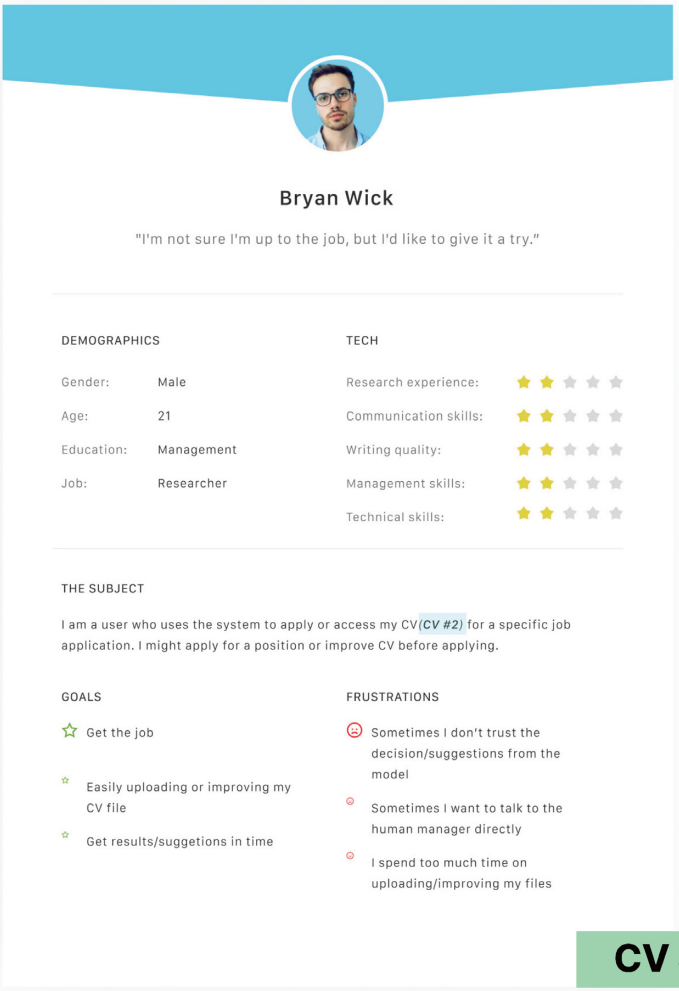
Poster



Target user



CV #1



CV #2

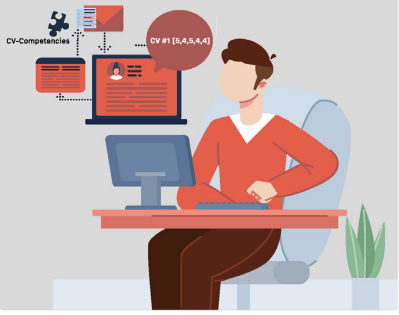
Storyboards for Task 2 & 3

Storyboard 1

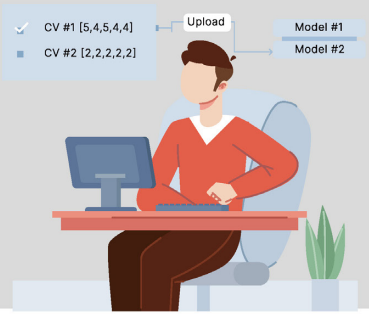
View this storyboard illustrating how an HR Manager engages with two AIs to address a dilemma when faced with difficulty in selecting a candidate.

1. The HR manager wants to know who would be a better candidate to get the job.

2. The HR manager uploads CV #1 to the CV-Competencies model for a 5-skill rating.

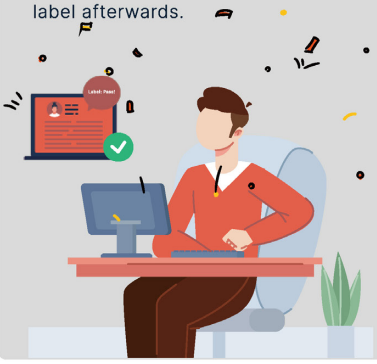


3. After reviewing CV #1's scores, the manager uploads them to the CV-Screening model for a recommendation.



4. The CV-Screening model gives positive or negative suggestions based on those points.

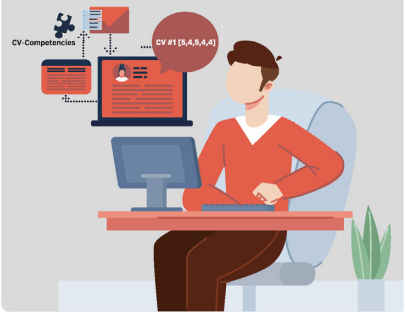
5. The HR Manager is satisfied with the label afterwards.



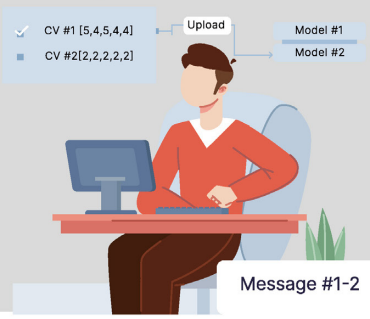
Storyboard 2

1. The HR manager wants to know who would be a better candidate to get the job.

2. The HR manager uploads CV #1 to the CV-Competencies model for a 5-skill rating.



3. After reviewing CV #1's scores, the manager uploads them to the CV-Screening model for a recommendation.

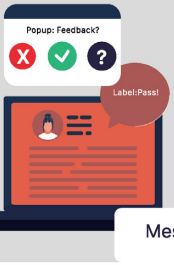


Message #1-2

4. The CV-Screening model gives positive or negative suggestions based on those points.

Message #3

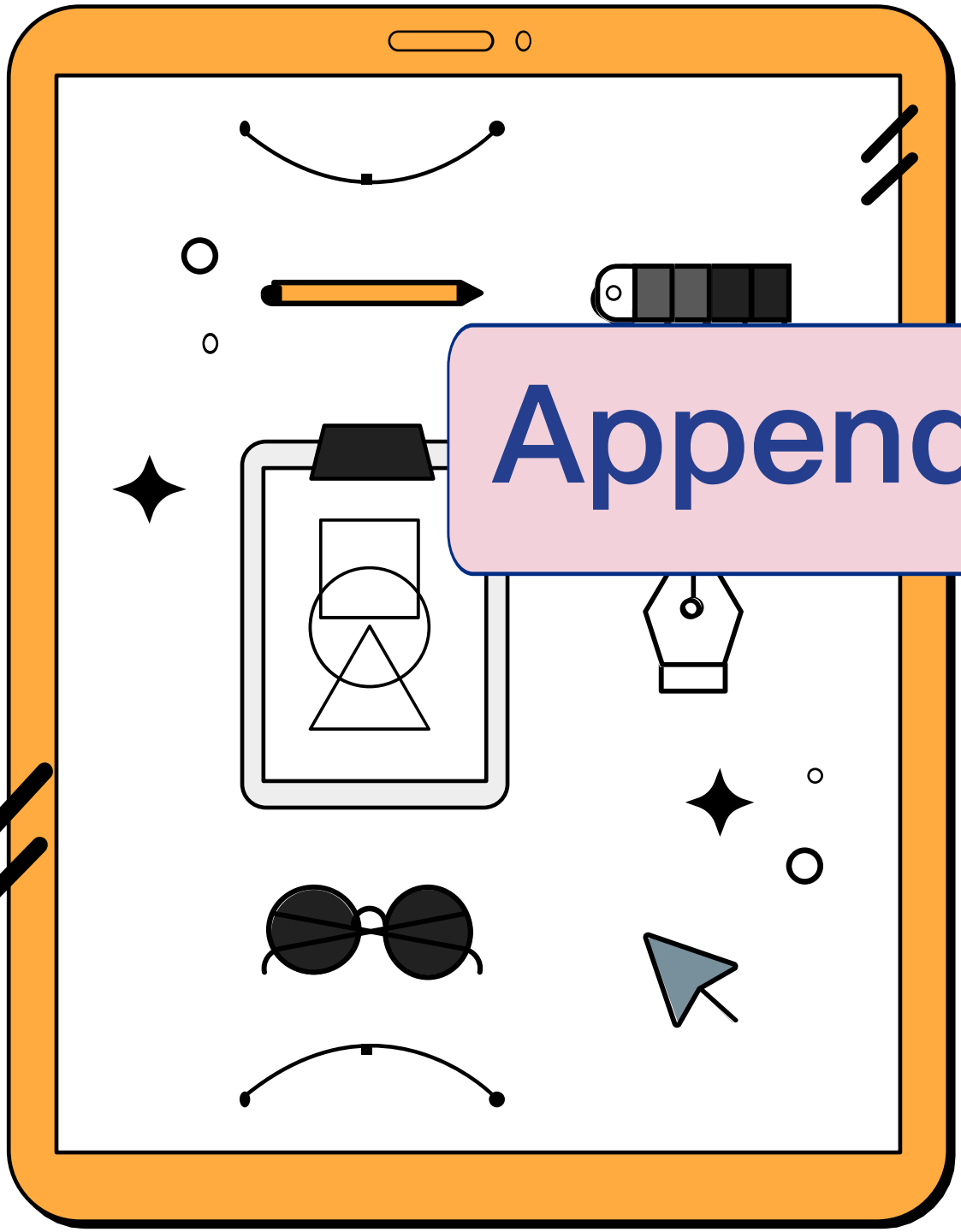
5: The CV-Screening model expects the HR manager to provide feedback on the suggestion given.



Message #4

6. The HR Manager is satisfied with the label afterwards.

Message #5



Appendix K

Transcription for Participant 1

Pre interview

Jiayi Zhou

Participants

"How long do you study for design?"
More than 5 years.

"What kinds of tools do you always use for the design process, like persona, user journey map, empathy map and so on?"
Yes, I always use persona and the user journey map to help me empathize and ideate during designing.

"Do you have experience in design products related to AI?"
No. I learned something about it on ITD, but for the final design in that course, we didn't work a lot on AI.

"What kind of digital tools or products have you used to help you design interaction between humans and AI?"
I don't have any idea. Maybe the new version Figma? It has something related to Chatgpt.

Message coding makes it difficult to understand the concept as a whole.

Task 1

(Silence for a long time...) "Could you think aloud? I see you are struggling." Well, I am curious about what Message is. There are a lot of visualized elements. I think I don't get it. See Figure 1.1. Could you give me some cues?

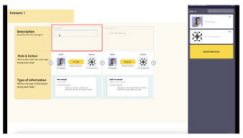


Figure 1.1.1 Message

I think here is the message and you see after clicking I need to type something. And the question is what the Message is. A little confusing for me. (After Jiayi explained what Message means in the project again) "So now what the Message #1 is?" I don't know... Maybe Message #1 means how the HR Manager gives the AI CV #1 to read. And Message #2 shows a process of how AI assesses CV #1. And I could find the responding steps in the storyboard. See Figure 1.1.2.

Figure 1.1.2 Explanation

Hard to find functions they need

Too many texts to read

Where is it (Prediction probability)? I need to find it by myself? OK, I think I find it. Here, right?

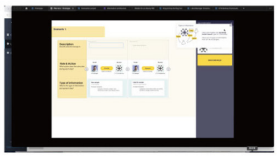


Figure 1.1.3 Hard to find

Change Model #1's function? So it is to change its model. Oh, so many choices! Let me try to click it. It works! Done.

"Well done! Is there anything you want to share with me? Suggestions or others." Yes, I spend some time on understanding what Message is although you told me before tasks. It is the examples listed on the Type of information helps me. See Figure 1.1.4.

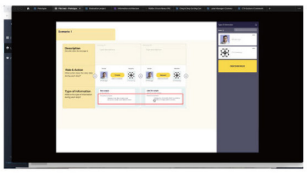


Figure 1.1.4 Examples

Classify the roles

And for roles, see Figure 1.1.5, I also spend some time on understanding what roles are. There is no cue for me to know if the role is a human or a model.

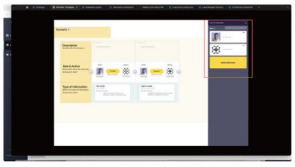


Figure 1.1.5 Roles

Don't understand the content of message

Misunderstanding on the task's description

Task 2

For the interaction to change the order, I think it is logical. It is the same as other tools.And I could understand what these messages are now. But I am curious about these two steps. See Figure 1.2.1. These two steps seem to have the same descriptions. For the Message #2, I think the AI model only provides the negative or positive feedback. That means the result would be "Yes" or "No". However，in Message #3, the model would give more details about the last positive or negative labels.

For Figure 1.2.1, I don't know if there would be any interaction between the HR and the model. I know the HR now is satisfied with the model's result, but I am not sure if the HR should give it a feedback or not. "After reading the related type of information, do you think so?" Well, after reading the type of information, I understand there is one feedback which is provided by the HR.

I have one suggestion. For the first time I saw the yellow button "Provide", I think it is a button. But then I find I cannot click it. "It is a button, but for the Task 2, it is fixed. Because if you could click the button, then the information would be changed. To keep there is only one information, the button is fixed. In the Task 3, you could click it." Get it!

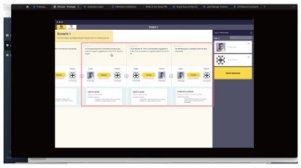


Figure 1.2.1 Message meaning

The information architecture should be changed on the right part.

Visualize the direction instead of texts

More attractive animations

And I don't see the direction between the sender and the receiver. That means I don't get the information "Who is providing the information with the other role?". Maybe it would be better if there is one arrow to show the direction.

For creating a new blank message, it is easy for me!

For the Touchpoint, let me see... First, when I click the first message, something jumps into the interface. I see a big space for the Type of information... What if I click the type of information? Nothing happens. Well, I see. I need to click the line on the bottom.

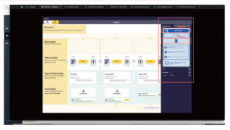


Figure 1.2.2 Touchpoint

For Figure 1.2.3, I need to change the label. Ok, I think I finish it. Wait, why there is nothing changed? Oh, I have to confirm it!

Figure 1.2.3 Touchpoint

More attractive animations

"Could you share your experience with Touchpoint?" Yes it is the information the humans would communicate with the AI model, right? "Does it help you understand how to create the human-AI interactions?" And there is something I want to share. You see in Figure 1.2.4, the purple color is so strong that I cannot pay any attention to others. However, the purple part is not the most important information. The most important one is on the bottom. I think the information architecture is a little strange.

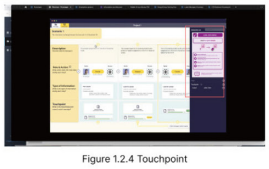


Figure 1.2.4 Touchpoint

Show the most possible type of information on the top

Visualize the types of information

Too small to read

Task 3

It is easy to choose roles for Sender and Receiver, but it confused me when I have to choose the type of information. See Figure 1.3.1. I don't know which one I could choose. The terms are abstract. "You could find more details about each type of information." Oh yes, I get it.

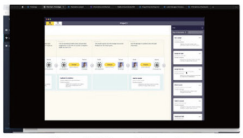


Figure 1.3.1 Type of information

Well, where do I change the style? Oh, I see. It is too small to read and click!! See figure 1.3.2.

Figure 1.3.2 Touchpoint

Transcription for Participant 1

And the most annoying design for me is like Figure 1.3.4. I spent a lot of time on exploring if there is anything I could interact in the colorful part. There is no cue saying it is one part related to the information instead of the touchpoint.

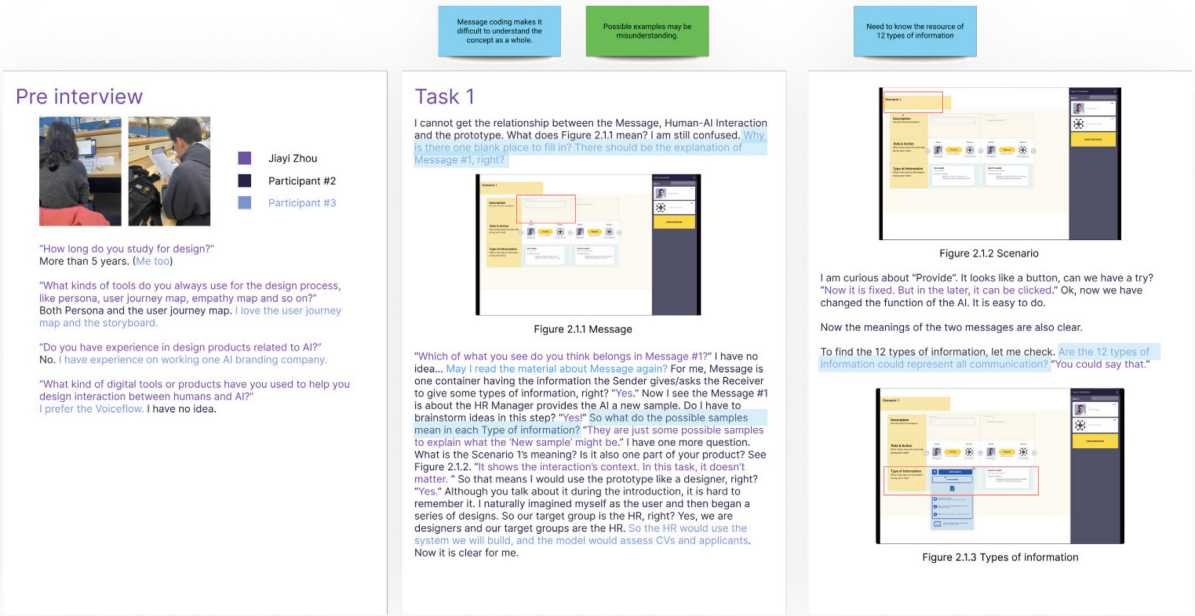
Figure 1.3.4 Touchpoint

And the other part is about the "Provide/Request". See Figure 1.3.4. I think maybe it could be designed as one arrow. It is more visualized and easy to understand. Now "Provide/Request" is confusing for me. I cannot get cues about who Provides the type of information with the other one. See Figure 1.3.5.

Figure 1.3.5 Sender/Receiver

And I get why you asked me what kind of design tools I would use during designing. This looks like a user journey map, and perhaps my familiarity with this tool will also help me subconsciously understand quickly what it takes to make sense of this.

Transcription for Participant 2&3



Transcription for Participant 2&3

Touchpoint means the UI for each message? The touchpoint of each message?

(After changing the CV #1) Is it successful? I am not sure... I think we did it.

I found Task 2 to be similar to a modification task. May I ask if this is the formalized workflow in the prototype? Is it that after I create the message sequences based on the use case, the prototype automatically generates these Touchpoints for me and I modify them if I am not satisfied? "No, it is just one task. The touchpoints are content created by designers." I think Popup should be the UI element, but it is Feedback.

Figure 2.2.2 Changes

What does the Loop mean? "You could click the Message #4." Oh I see. It is about where it would go back. Agree!

I was curious about the content in Touchpoint. Are there places for designers to choose some UI components for each Message? See Figure 2.2.3. "No. Its purpose is to help designers understand what data is being passed in each Message." I think they are too specific and have too many design details, like a finished UI interface. This may limit the designer's ideas.

Figure 2.2.3 Touchpoints

For me, designer brainstorming isn't just pie-in-the-sky rambling. We are thinking out of the box with some basis. Often, the touchpoint in the user journey map is a very good starting point. I'll use them as a starting point to think about more potential ideas. Simply put, I'll look at the touchpoints to find the actions that exist and design interactions that match them.

"After the Task 2, how do you think of Touchpoint? Does it help you understand how it works or does it help you understand Message?" To be honest, I am still confused until now. For me, touchpoints are the many results after brainstorming. They shouldn't be like one UI component in detail. There seems to be some reversal of direction now. Maybe in Task 3, after doing all things by ourselves, I could understand it.

The prototype has a lot of abstract concepts. It is hard to educate designers how to use it. Probably because I have extensive experience using the user journey map, I could quickly understand how to use this archetype in the first two tasks.

Task 3

I click the Receiver, but nothing happened... Maybe we could click the HR manager here to fill it in. Ok.

Figure 2.3.1 Fill in roles

I think all types of information could be more visualized.

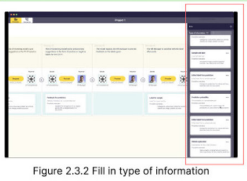
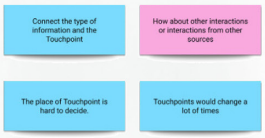


Figure 2.3.2 Fill in type of information



I hope the explanation could be bigger to read. Now it is too small. See Figure 2.3.3. Me too.

Figure 2.3.3 Too small to read

"Why do you choose Provide here?" See Figure 2.3.4. Oh it is Request! We forget it! And here is one Confirm button, right? It is too easy to miss it! Maybe you could give it one attractive and different color for Provide or Request.

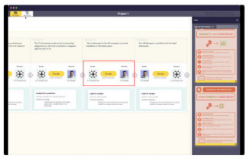


Figure 2.3.4 Provide

For the touchpoint, I have no idea how to make it better. The question is similar to "which came first, the chicken or the egg". Sometimes we decide on Touchpoints first, sometimes we decide on something else. For example, Upload is a touchpoint that we don't even need to brainstorm, it's needed whenever content needs to be uploaded.

Is there any example including complex touchpoints? I think touchpoints are something that designers need to keep revising, not something that can be decided in one go. I am still skeptical about the part.

And I think we need to think about senior designers. They have rich experience in designing and they could have a lot of Design Intuition. For them, maybe they don't need the modules in Touchpoints. In one word, I still think it is a little strange. What's more, there is less connection between AI and this part. But they show the instances related to the types of information. Well, maybe it could be designed like Figure 2.3.5.

Figure 2.3.5 Touchpoints

Does the corresponding UI component exist for every content passed in a Message? Could there be something outside the system, such as a text message, an email? It's something that jumps off the interface. Will they still exist in UI components? "Good question." I hope we could pay more attention to Human and AI, Touchpoints are too general. Or they could be added into the Type of information part.

After

Does the yellow button plus the Loop? I think so. See Figure 2.4.1.

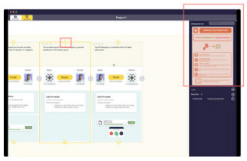


Figure 2.4.1 Yellow plus

What's more, how do designers know what kind of touchpoints they could choose? Is it related to Type of information? "Yes." But what if designers want to do something out of the Type of information? See Figure 2.4.1. "What do you mean?" For example, for the Feedback for prediction, I don't want two buttons like "Yes/NO". I want to do one form or a reliability scale. What should I do? "It is related to the UI component in detail, and out of the usage of the prototype." It is a little strange if I have some ideas but I cannot write or draw it down here, at least for me. Maybe here is one place for designers to write or draw freely. And if the part is trying to inspire designers, it has too many instances and UI components. It will impact designers later.

Go back to the two kinds of buttons. I still think they are confusing. One visualized icon for one function. Secondly, it is linear. I have some more ideas. When you just give designers one linear structure, they are hard to think about loops. But if you tell them, they could also build loops, there will be a lot of loops. Maybe it would be better to do a modular content. See Figure 2.4.2. So designers could put them like one empathy map or stuff like that. Or keep the linear part, and give them more places to change structures, too.

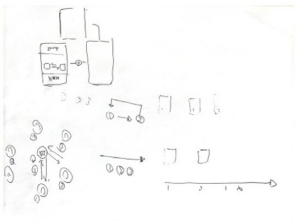


Figure 2.4.2 Message

I have to say that content about characters, actions and types of information is very important. It would be best to make it more prominent.

Transcription for Participant 4

Message coding makes it difficult to understand the concept as a whole.

Show Message directly might be better to learn

Icons have different meanings for different designers

Hard to find Type of information

Pre interview

Jiayi Zhou

Participant #4

"How long do you study for design?"
More than 5 years.

"What kinds of tools do you always use for the design process, like persona, user journey map, empathy map and so on?"
Both Persona and the user journey map.

"Do you have experience in design products related to AI?"
No.

"What kind of digital tools or products have you used to help you design interaction between humans and AI?"
Voiceflow, Posenet, TensorFlow.

Task 1

Ok, I see the two examples. Now what should I do? Task 1...Where is the Message #1? See Figure 3.1.1. It is one blank space.

Figure 3.1.1 Message

If I have to fill it in, let me see... Well I think I don't understand the basic concepts although you introduced them. Ok... I think here is the Message #1. See Figure 3.1.2.

Figure 3.1.2 Message #1

One question. Who could see this interface? "Designers." Ok... Why are the Sender and Receiver like this here? Is the relationship between these two mutual? "Yes. But in this message, only the HR is the sender." Ok. Why is there one plus button? See Figure 3.1.3.

Figure 3.1.3 button

"What do you think about it?" It would connect to the other message, for me. Now I think it is to plus one blank new message. Well, now the Message #1 is to upload. The Message #2 is also in the same step.

To find the "Prediction probability"... is it in Figure 3.1.4?

Figure 3.1.4 Type of information

The task is not so realistic

More animations

A little hard to know the Fail/ Pass

Request/Provide is easy to miss

More visualization on how to pass it

Need to know the resource of 12 types of information

The early stage is in mess

Touchpoints have too many details in UI

It is easy to find. But I am confused about 12 types of information. Are these types of information something recorded by the prototype? I mean, something like a chat history. "No, they are basic concepts for all use cases." OK.For me, it may be hard for me to decide what kind of information would be exchanged in this step. Because now all data are a little too abstract.

Task 2

I don't understand why there are repeated information in the Roles & Actions. What function is this part? What do they want to show me? See Figure 3.2.1.

Figure 3.2.1 Role & Action

"They are different. And the whole message should include one Sender and one Receiver, and types of information. If we only see the Roles & Actions, we lose the information they exchanged." Get it. And I am confused about the Action. See Figure 3.2.1. I don't know who sends the information. Oh I see, there is one text called "Sender".

I think I need to change the order. Great, easy to finish!

Now I could go to the Interaction version. Easy to do it. Wow, what it is! Ok, now I have to correct the content. Is it successful? See Figure 3.2.2. I am not sure. It seems I did it successfully.

Now I have to change the Fail into Pass. What is it? Fail means the manager doesn't agree with the AI's result and hopes it to work again. So in the final message, the manager needs to provide new information and then the model would work.

"The fail is the label of the model. " So that means the Model works in a wrong way and the manager has to correct it? "Oh no, it is just one way for you to experience how the Touchpoint works. " Get it. But I think it is not realistic. In most cases, the HR would not pay attention to CV's prediction "failed". He or she would just delete it.

Figure 3.2.2 Missing

"Is there anything out of your expectation?" I am curious about the Loop. Is it fixed or shown by the prototype automatically? "No, it is also designers's work." Ok.

I think designers would be more familiar with the content in Touchpoint, and that really helps me to understand what happens in each message. Why do you put it on the bottom?

But I think it is too strange on the content of Touchpoint. During the early stage of ideation part, it is hard for designers to think about so beautiful design components. Now it is more like to assess one current UI components related to the functions. It is not like the messy and crowd things we always use for brainstorming. We have to remember in this stage, we don't have prototypes. Maybe it would be better to look like something in wireframes.

Visualize different types of information

Less description

Immersive in message

Need too much logical cues

Touchpoints are too detailed

More freedom to create in the Touchpoints

The task is a little hard to understand

The same interaction way for the same purpose

Request/Provide is easy to miss

Task 3

For Figure 3.3.1, I think the new step should be added here. Oh , nothing happens. "Why do you want to do that?" Because the machine wants the HR to provide something, then the HR does it. So in total, there would be two steps.

Figure 3.3.1 Addiation

Why is there nothing when I click the Receiver? Oh, the same function part as the Sender, right? See Figure 3.3.2.

Figure 3.3.2 Receiver

Oh, I forget to change the Provide/Request.

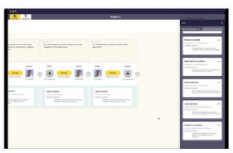


Figure 3.3.3 Change buttons

After

I think designers should have some background about the technology before they know how to use the prototype.

What's more, I think it is a little hard for designers to give out some logical sequences in the early stage of ideation. Sometimes I am not sure what would be the design output and what the information would be exchanged in each step.

And for the Touchpoint, it is now the basic and classic UI components. You know, there are many other options to solve the same question or meet the same needs. If you give on example here, it might limit designers imagination.

But there are two benefits I love much. First, this tool helps me to sort out how HR's needs are to be matched with AI and how his tasks are to be buried at each step.

Also, I like designs that hide functional areas as much as possible. This lets me focus more on creating relationships between Messages. This also has a greater view and space.

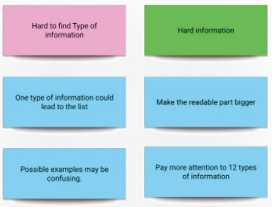
And I have some suggestions on the UI in the prototype. First, it is hard to see the difference between types of information. Maybe we could use colors or some icons to make it different. And it could help designers to spend less time reading texts.

Secondly, it is the button for Provide/Request. It is easy to miss now.

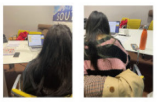
Finally, the description could be more simple. Maybe one word to conclude.

For touchpoint, maybe designers want to use the moodboard or test to show what they want. It would inspire them more.

Transcription for Participant 5 & 6



Pre interview



Jiayi Zhou
Participant #5
Participant #6

"How long do you study for design?"
More than 5 years. (0.5 year)

"What kinds of tools do you always use for the design process, like persona, user journey map, empathy map and so on?"
Both Persona and the user journey map. I love the user journey map and the storyboard.

"Do you have experience in design products related to AI?"
No. Me neither.

"What kind of digital tools or products have you used to help you design interaction between humans and AI?"
Posenet. I have no idea.

Task 1

For the Message #1, I think it should be like Figure 4.1.1.

Figure 4.1.1 Message #1

For 12 types of information, why are there some examples? What do they mean? See Figure 4.1.2.

Figure 4.1.2 Examples

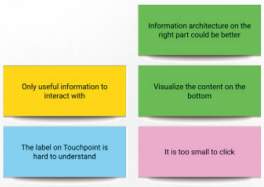
Oh! I know. They are just examples for what the new sample could be!

12 types of information are the possible examples? No... Oh I can click it. But it is too small to read the texts. For me, the name of the term "Prediction probability" shows some connections with the sample. Maybe we could find it from the possible examples? Let me try... I have no idea and try to click everywhere I can to see what will happen. Oh, I find it! It is really hard to find it. And for me, for Figure 4.1.3, the blue words are hard to read.

Figure 4.1.3 Blue words

Is the content on the right the manual for the interface? No. It is some database for designers to drag and drop. Well, if that is one database, I think it should be put on the left. We always pay more attention to content on the left. And is it something users created? "No, it is the prototype has by itself and users cannot change it. " Ok... I think in that way, it is ok to put in on the right. How do we could go back to the Roles? Oh! I get it. It is a little hard for me to do that.

Figure 4.1.4 Hard to find



I see, you want to implement the "collapse" animation, right? That makes sense. You can modify the animation. Or you can take a page from Ps or Adobe's tabbed design, where they arrange all the features together.

See Figure 4.1.5, do I finished it? Oh, I think so.

Figure 4.1.5 Missing

And I think "Edited" is not good. Maybe another word.

Is Message important in this prototype? "Yes." What is its function? "It shows the information exchanged between human and ai. And one message sequence is the Human-AI Interaction." Ok, get it. It is really complex. For me, designers are responsible for the front end, like the UI style and the workflows. Now these material are more like something for software engineers. I agree. For designers, these concepts are not so difficult but they need more sensing timing. More information and explanations for them to read about the message and types of information.

Task 2

We need to change the order...Yes, easy for me. It is very logical and simple.

For deleting one message, maybe one icon like Reducing is better. Now Let's go to the Interaction version. Easy.

For deleting one message, maybe one icon like Reducing is better. Now Let's go to the Interaction version. Easy.

Change the CV #2 into CV #1.. Oh it is too small to see! See Figure 4.2.1 Because of the architecture and the color in the right part, I would pay more attention to the blue card instead of the information on the bottom. The color part works as the explanation. I prefer to put it on the bottom or put it in another information architecture. Like one button or one icon, then the designer could click the button to see the colorful card.

Figure 4.2.1 Small

I don't get what fail means. For me, it means the CV uploads unsuccessfully. It is strange. You see here is one green 100%, and one read fail. It is contradictory.

Figure 4.2.2 Fail

After reading the description, I know the Fail means the label. My suggestion is to put it on the other place. Or delete the green 100%. Maybe we could also use another word like "Out" or "Unqualified". "Fail" has two meanings, and it would make designers confused.

"Until now, what do you think of the Touchpoint?" They are the places for human to interact and give triggers for the AI to work. The same as the Touchpoint in the user journey map. A classic word in the user journey map.

The other suggestion is about the Touchpoints' choices. See Figure 4.2.3. It's so hard to tell the visuals here. Would like it to be more visual if possible. For example, "Pass" is green. What does the Metric mean? "It is related to another type of information, like Prediction probability. " And if there are two blanks for changing, we could use some animations to hide some information in the beginning. And if in this interface we cannot use the Metric, let it be grey or disappear. I think the amount of information is not so much, just keep it directly.

Transcription for Participant 5 & 6

Figure 4.2.3 Touchpoint

I have one question. If there are two human roles hoping to communicate with the AI, what should designers do? Create one more message? But how about the order? And What if the human role wants to exchange one more information at the same time?

By the way, I am curious if it has connection with the user journey map. It gives me the same sense. I agree, especially the word "Touchpoint". That reminds me of the user journey map and try to understand what the whole sequence is based on my knowledge about the user journey map.

Task 3

Well, when I click the Receiver part, nothing happens. Should we click the HR from the right part directly/ Ok, it works. See Figure 4.3.1.

Figure 4.3.1 Roles

Maybe it is better to have the same interaction. I agree. "How do you think of Provide or Request here?" For me, I forget to switch it as the Request. Wait...I am not sure...Because before the AI requests the HR to do something, it should provide some ways or visualized signs. So it might be better to add one more step before that. You mean it is not one new information, right? Yes.

For the Touchpoint, they are too small to click. See Figure 4.3.2.

I have one question. See Figure 4.3.3. These UI components seem to be clickable. Can designers click them? "No." How about the HR? "No. They are working as inspirations during the brainstorming, and the interface in detail should be built in another tool like Figma." What does the final black button mean? "Ask for more explanations." They seem to be in different information architecture. You could give different visual icons and layers for these information.

Figure 4.3.2 Touchpoint 1

Figure 4.3.3 Touchpoint 2

I am curious if the Touchpoint really helps designers instead of bringing limitations for designers. Now you show me some UI components and some buttons for interactions, and it would impress me when I use figma to create UI components later. However, if I could have choices, like choose one from one database, that might be better to free my mind.

I have one suggestion on the Provide/Request. Maybe it would be better to show the direction. I see here is one Sender and One Receiver, see Figure 4.3.4. But it is not so visual. Maybe some icons could be better.

Figure 4.3.4 Visualization

I have some suggestions for visual elements, too. See Figure 4.3.4. It's ok to keep names of roles, but Provide/Request could have some different colors. Then designers could get all information in general even thought they just have a second to look at it.

And I don't know how the types of information work. Who could classify these types? It needs more background. And some words like Feedback, it is a little different from what I think.

So the type of information is important, right? How about combine the second and the third lines? I mean designers could see roles in one line, provide/request types of information in the next line. That might be more clear. And add the instance to explain what the new sample is. So that is the Touchpoint, right? "Yes." Well, now I get the idea better. It is better to show touchpoint first, for me. The order to show information is important.

Firstly, I could see less abstract items. Then it is the abstract type of information. So change the order of the two lines. Designers would get the touchpoints much easier than the Message. Or use some icons to visual the type of information. See Figure 4.3.5.

Figure 4.3.5 Visualization