FABIO AMELIA

27 Jan 25 - 31 Mar 25

Human-Computer Interaction for AI Systems Design

Final project assignment

UNIVERSITY OF CAMBRIDGE
ONLINE

Table of content

The task

Consider the design of an AI-assisted decision-making tool for a manufacturing plant. The plant uses machinery to produce heating control units for homes and unfortunately, this machinery tends to fail, sometimes at inopportune times. When a machine fails a replacement machine must be used. However, such a replacement machine requires workers with different skills than the usual machine. So, when machinery fails there is a need to quickly (1) identify a replacement machine; (2) reroute skilled workers to the new machine; and (3) ensure manufacturing commences smoothly using the replacement machine.

The AI-assisted decision-making tool is implemented as an optical see-through head-mounted display, allowing the user to obtain information about each machine, such as the types of skills required to operate it and its range of capabilities, by merely looking at it on the shop floor. The tool is meant to be used by a supervisor working in direct contact with workers and machines in the plant, providing advice on which machine to use and which workers to reroute based on their workload and expertise.

All of this project is my own work – none of the text was written by AI. Since English is not my first language, I acknowledge that I used Grammarly to identify some grammatical errors and help me correct them.

Chapter 1
Solution-neutral problem statement and requirements

To reach a solution-neutral problem statement I made a design problem analysis regarding the objectives, constraints and design space.

Objectives

- To effectively manage the manufacturing plant during a crisis.
- Directing the right kind of workers with different skill sets and expertise.
- Manage the different types of machines.
- Minimise downtime when the machines stop working.

Constraints

- Head-mounted display size and weight.
- The battery life needs to last at least 8 hours.
- Workers with different skill sets need to be managed.
- Manufacturing plant physical space.
- Replacement machines with different functionalities.

Design Space

- Machine learning to identify defective machines.
- Machine learning to recommend the appropriate replacement machine.
- Machine learning to associate the worker with the right skills with the new machine.
- See-through head-mounted displays that are effectively portable to be worn during the whole working day.
- Headset's UI for the effective interaction of the manager with the AI assistant.

Solution-neutral problem statement

"Design a system that identifies failed machines and helps manage workers and replacement machines during a crisis."

Reasoning

The problem statement obtained is the most solution-neutral while still providing enough context for the environment and the constraints. It clearly states the necessity to manage both the workers and the replacement machines. I removed the reference to the AI assistant because that was already giving a solution to the main problem.

Chapter 2
Function model and morphological chart

Function model

I started by designing a simple level function model.

This helped me envision the system's most important inputs and outputs.

I then expanded it into a decomposed level where I identified four key functions.

Function 1: Collect observed data

Function 2: Identify problem

Function 3: Identify appropriate replacement machine

Function 4: Identify worker with right skills

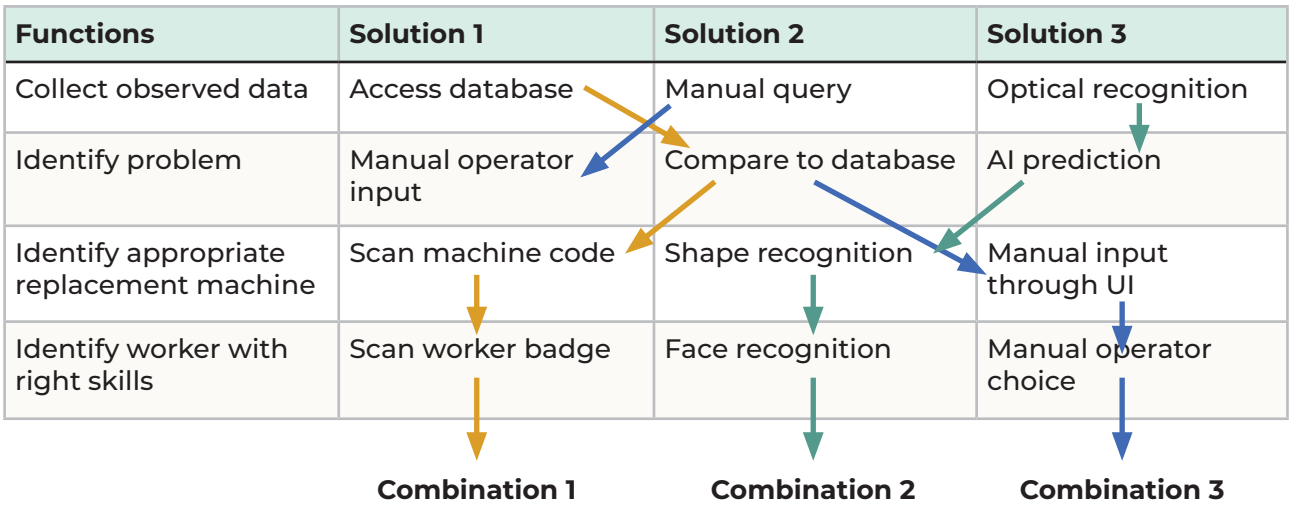
Morphological chart

I created a morphological chart with the four functions, and I found three possible solutions for each of them.

Functions	Solution 1	Solution 2	Solution 3
Collect observed data	Access database	Manual query	Optical recognition
Identify problem	Manual operator input	Compare to database	AI prediction
Identify appropriate replacement machine	Scan machine code	Shape recognition	Manual input through UI
Identify worker with right skills	Scan worker badge	Face recognition	Manual operator choice

Morphological chart.

I then traced lines to create three different combinations.



Morphological chart with combinations.

Combination 1 is more focused on a database-centred solution. Combination 2 uses AI prediction and optical recognition the most. In Combination 3 the user will input data manually through the UI.

Conceptual designs

I insert my three combinations into a concept evaluation table.

The chosen criteria are “Speed”, “Reliability”, and “Interactivity”. Speed and reliability are the most important factors during a crisis.

I decided to carry over combination number 2. This solution can use more advanced technology and after good training can perform better in terms of speed and interaction.

Brief narrative

The plant manager will wear an optical see-through head-mounted display that will indicate the issue in the machine, the AI predicts the problem, suggests a replacement machine thanks to shape recognition and indicates the right worker thanks to face recognition.

Chapter 3
Automation strategy

I analysed the functions, skipping the number one because it cannot be automated.

Function 2

This function can be automated by making the head display highlight what exactly is not working correctly, and then the AI can inform the manager.

- **Automation Type:** Analysis automation. The system will have to analyse the collected data and determine the issue.
- **Automation Level:** 9. The system will capture all the images and only inform the user in case there is something to do.

Primary evaluation criteria

- **Mental workload:** Improved. The manager will not have to analyse where the problem is anymore.
- **Complacency:** High. The manager will not be able to detect the defective machines.

Secondary evaluation criteria

- **Automation reliability:** High. In case of a false positive, the working environment could be disrupted.
- **Cost of action outcomes:** Very high. The plant could stop working for a long time.

Function 3

The AI will access the database and decide what is the right replacement to be used.

- **Automation Type:** Decision automation. The system will suggest what replacement machine to use.
- **Automation Level:** 4. The system will suggest one alternative machine.

Primary evaluation criteria

- **Mental workload:** Improved. The manager will not have to decide on the right replacement machine.
- **Situation awareness:** Improved. The manager will get extra help and context.
- **Complacency:** High. The manager will not be able to detect the replacement machines alone.
- **Skill degradation:** High. If the system cannot direct the manager to the right replacement machine, he will not be able to do it alone.

Secondary evaluation criteria

- **Automation reliability:** Low. In case the AI choose the wrong machine, the manager should still be able to choose alone.
- **Cost of action outcomes:** Very high. The plant could stop working for a long time.

Function 4

The AI can suggest which worker is the most suitable to work with the replacement machine.

- **Automation Type:** Decision automation. The system will suggest which workers are the most suitable.
- **Automation Level:** 3. The system will suggest a selection of workers down to a few.

Primary evaluation criteria

- **Mental workload:** Improved. The manager will not have to decide which are the right workers.
- **Situation awareness:** Improved. The manager will get extra help and context.
- **Complacency:** High. The manager will forget how to detect the workers alone.
- **Skill degradation:** High. If the system is not able to direct the manager to the right workers, he will not be able to do it alone.

Secondary evaluation criteria

- **Automation reliability:** Low. If the worker cannot work on the replacement machine, it can always be communicated.
- **Cost of action outcomes:** Very high. The plant could stop working for a long time.

Chapter 4
Interaction strategy

I enhanced the functions thanks to the mixed-initiative interface principles.

Function 1

The manager will point to the machine, the AI will make it glow, and the user will confirm the selection with the fingers or by nodding the head.

Utility: Augmentation of the manager's capability.

Balance: This is a special collection mode that the user triggers when required.

Function 2

The user will receive a message. In case the AI is identifying a false problem, intervention through spoken means is possible.

Balance: This is very important. False alarms will slow down the factory.

Control: The manager will be able to manually identify the issue.

Function 3

The suggestion of the appropriate replacement machine is of crucial importance. If the manager can identify the correct replacement, the factory can go back to full operation very soon.

Balance: The cost of interrupting the user at this stage is very low because the manager is already alerted.

Function 4

This function can be displayed on the manager headset.

Control: The manager will decide who works with the machine.

Uncertainty: The manager knows that the AI could suggest solutions that cannot be applied in reality.

The human-AI teaming issues that arise are:

Alignment: If the AI gives a series of wrong suggestions, this will deteriorate trust and alignment. There is the risk that the manager will stop using the AI.

Interpretability and trust: Does not apply. The manager is a senior user who knows the plant and the workers already. The AI can give suggestions, but the final decisions are still under the management's responsibility.

Managing AI model update: Does not apply. The data the AI is trained on are those of the machines and the workers that are operating in the plant.

Chapter 5
Interpretability report

The nature of the data the system is processing

Function 2

The reasons why a machine is not working properly can be several, known and unknown issues. There is a high uncertainty, and the diagnosis could rapidly change.

Function 3 and 4

The data is not going to change rapidly. The machines and the workers are carefully monitored and kept in a database.

Information visualisation methods

To identify the machine problem, I will use a heat map overlaying the machine, highlighting the part of it that is not working properly.

A Radar chart will represent both Workers and Machines. For the workers, I will display type, seniority, expertise, and company level. For the machines, I will display the required expertise, type of worker, wear, and state. Comparing these two charts, the manager will decide which worker is appropriate.

Potential interpretability problems

Function 2

This function has Trade-off problems. The manager will want to know instantly if the machine is faulty. If the AI is not correct, this will drastically affect the plant. The AI must analyse the data correctly, and this could also slow down the operations.

Function 3

If the manager is strongly opinionated on a solution and the AI suggests the same, the manager would not consider other paths, bringing Confirmation bias.

Function 4

Safety, Fairness, and Alignment need to be taken into consideration.

Methods to enable interpretability

Function 2

I will use a Confidence Score. In case the score is high, the manager can decide to replace the machine immediately.

Function 3

Explaining the Factors is of great use. The manager will trust the AI more if he knows which data motivated the AI to make the decision.

Function 4

A Confidence Score could be useful.

Chapter 6
Sharing of control and user agency

The manager can control the AI in different ways:

- While working, the manager will receive alerts highlighting what machine is not working. The manager can ask for recommendations. This is a **Strategic** shared control level because the manager asks the AI for a very high-level recommendation.
- When the AI suggests a list of replacement machines or workers to operate the machine, the manager can operate a **Tactical** shared control level because alternative suggestions based on feedback can be requested.
- When the manager has decided what replacement machine and worker to use, the **Execution** shared control level will be used as the manager will then do a low-level execution by telling the AI the suggestions are ok and he is proceeding to use it.

Ownership

The manager has a strong ownership over the system. In case the AI detects a malfunctioning machine, the manager can decide to start a manual inspection. When the AI suggests a replacement machine or worker, the manager can still decide to make a different choice if it is more adequate. The manager needs to keep a feeling of agency all the time to guarantee smooth operations.

Prompting automation

A form of prompting through spoken means is required. Due to the emergence of the machine malfunction and the speed required to detect a replacement, it is mostly unlikely that the manager will have time to use a keyboard to prompt instructions to the AI. This can be solved by using direct speech recognition. The AI will detect the manager talking and will translate the speech into written form to use as a prompt.

Risk of inaccuracy

There is always the risk of inaccuracy. The AI could wrongly detect the malfunctioning machine, or suggest a worker that is not able to operate the replacement machine. This can be improved by using interactive machine learning before the usage of the AI, but it could still be conducted during working hours.

Solution principles:

- **Provide effective data representations:** The user needs to be aware of the results while operating.
- **Exploit interactivity and promote rich interactions:** It is important that the user can interact with the AI as desired because there could be different interactions that were not taken into consideration during the machine learning phase.
- **Engage the user:** The designer needs to consider the user's need to feel engaged with the testing and the learning process.

Chapter 7
System boundary and risk analysis

System boundary

Included in the boundaries

- The AI system
- The network connectivity
- The manager
- The machines
- The replacement machines
- The workers
- The database with the machines' and the worker's data
- The headset specifics
- The augmented reality

Excluded from the boundaries

- The factory's policies
- The supply chain
- The maintenance
- The repair service

Keeping in consideration elements that are both hardware like the headset itself and the machines, software elements like the AI system or augmented reality, and also human elements like the manager and the workers, are all key factors to the system boundary.

What is not required to be taken into consideration are the supply chain's problems. The users still need to be safe even if there are problems in those regards.

System mapping

To map out my automation function I use a Process diagram.

With this diagram, all the interactions between the AI system, the manager, the replacement machine and the workers are mapped. When something goes wrong, the manager will request an alternative solution. The system will enter a loop until the AI provides a viable solution.

Risk assessment

I use the Structured What-If Technique (SWIFT) risk assessment method to calculate an acceptable level of risk.

ID	What-if question	Hazards and risks	Relevant controls	Risk ranking	Action notes
1	“Is there no network?”	Alarm not received by the manager.	Alert to notify the manager that the system is offline.	5	The manager needs to contact the Helpdesk to reestablish the connection.
2	“Is the system sending a false alarm?”	The system is sending false allarms.	The manager needs to notify the AI of the false alarm to further train it.	1	More interactive machine learning and machine teaching sessions are required.
3	“Is the replacement machine not available in the warehouse?”	The suggested replacement machine is not available.	The manager needs to verify personally.	3	The manager will have to request a new suggestion.
4	“Is the worker not available to operate?”	The suggested worker is not available.	The manager needs to verify personally.	4	The manager will have to request a new suggestion.
5	“Does the AI not find any valid solution?”	The plant could stop working.	The manager will receive a message stating that no good solution has been found by the system.	2	The manager needs to manually find a replacement machine and the worker.

Number 2 Is the most risky. In case the system sends a false alarm, the manager can stop the plant, causing money loss, or mobilise the repair service to do a job that was not required.

Redesign the automation function to reduce the level of risk

The system must be able to effectively communicate with the manager in case it is not sure of the suggested solution.

The AI needs to be constantly kept updated to know the workers' and the replacement machine's status.

Chapter 8
Verification cross-reference matrix

I insert my functions into a VCRM table.

ID	Requirement	Verification	Allocation	Success Criteria
1	AI must alert if offline	Demonstration	Wifi module	The lost connection must be highlighted on the interface
2	AI must alert if a defect is identified	Inspection	Headset interface	The AI successfully report a defect if identified
3	AI must suggest a replacement machine	Test	AI module	The AI is suggesting the right replacement machine
4	AI must suggest a replacement worker	Test	AI module	The AI is suggesting the right worker
5	AI must communicate uncertainty percentage	Demonstration	User Interface	The percentage is displayed correctly
6	AI must communicate a not up to date database and request update	Analysis	Database	A request for a new update is sent

Thanks to this table I can check that all the system requirements are met. The most important requirements are numbers 2, 3, 4. But I also included new requirements found in the previous chapter like numbers 5 and 6.

For numbers 3 and 4 I choose to apply a Test verification so that the system can injected with predefined inputs and check if the AI produces the desired output.

Validation strategy

To validate my entire strategy, I will use an analytical method by using the Human-AI heuristic guidelines.

Initially, the users will need to be aware of what the system can do, and where there is a risk of failing,

During the interaction, the UI will need to show contextually relevant information like the accuracy percentage.

When incorrect, the AI should easily adapt to the manager's request by facilitating efficient correction.

And over time, the AI will need to be able to update and adapt thanks to past interactions.

Deployment studies

After the system is successfully deployed, I will keep analysing it by using a combination of Logging and Surveys.

It is important to keep a log of all past interactions and user data to keep improving the dataset.

At the same time, also users' opinion is important. Thanks to focused surveys, I can correct the system by following the feedback obtained.And over time, the AI will need to be able to update and adapt thanks to past interactions.

Conclusions

Thanks to this project I could follow the creation of a Human-AI system from conception to final draft.

I followed steps like creating solution-neutral problem statements and requirements where I delineated the essentials of the system. Thanks to it, I could create a function model and morphological chart that deepens each specific function of the system.

I then used the functions I envisioned to define an automation strategy and an interaction strategy that helped refine how the user would interact with the AI. I used it to discuss interpretability issues that arose in the system and issues around the sharing of control and user agency that helped discover the limitations of both humans and AI.

Finally, I could determine the system boundary, an analysis of the key risks in the system and a final verification cross-reference matrix that helped focus the AI on only the most important aspect of the interaction.

Following all of this, I also created a validation strategy and a deployment study that envisioned how the system could be maintained in the future.

With all this data, I created this report, which describes a full AI system and how the users can interact with it in its totality.

Fabio Amelia
28 Mar 2025