UNIVERSITY OF CAMBRIDGE

HUMAN COMPUTER INTERACTION FOR

Ai SYSTEM DESIGN

Methods, principles and approaches for designing systems, products and services that leverage AI to assist users in achieving their goals and improving user experience.

OVERVIEW PROBLEM FUNCTION AUTOMATION INTERACTION INTERPRETABILITY CONTROL RISK VERIFICATION

AISHWARYA KARVE I CAMBRIDGE UNIVERSITY I FINAL REPORT

CAMBRIDGE ADVANCE ONLINE

OVERVIEW

what’s the brief?

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.

OVERVIEW

AI implementation

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.

OVERVIEW

design objectives

Update information regarding workers skill sets and availability
Swiftly replace the machine in case of failure
Determine which workers are skilled to use the replaced machine
Supervisor's requirements are met to avoid issues while transition
Minimise production downtime

Design Space:
AR head mounted display
AI assisted matches for assigning workers

⌛ USER
Busy multi-tasking supervisor

💼 BUSINESS
Impact on other business activities

🖨️ DATA
Inventory of workers skillsets and availability

🛠️ TECHNICAL
Accurate matching algorithm to ensure right workers are assigned

SOLUTION-NEUTRAL PROBLEM STATEMENT

“

Design a tool to minimise production downtime in case of machine failure for a manufacturing plant, considering workers' availability and skillset

”

Reasoning
I think this level of abstraction would help exploring all the options.
Research will give answers to what kind of technology and for what user group this tool would be helpful.

FUNCTION MODEL

a model that can optimise various processes, enhance productivity, reduce downtime, and improve overall efficiency.

Creating a function model and systematically investigating how to translate functions to function carriers, let us explore a wide variety of design options, helping us avoid missing out on great design opportunities.

FUNCTION MODEL OF A MANUFACTURING PLANT

POWER ON

MACHINERY DATA

WORKER DATA

POWER OFF

REPLACEMENT MACHINE (SUGGESTION)

ASSIGN WORKER (SUGGESTION)

MACHINE IDENTIFICATION & EVALUATION

MONITOR MACHINE PERFORMANCE

✅

❌ IDENTIFY FAILURE ❌

FIND REPLACEMENT

ANALYSE MACHINE DATA

🤖

MACHINE REQUIREMENTS

COMPARE

MACHINE AVAILIBILITY

MATCHING ALGORITHM

RECOMMENDATION ENGINE

ANALYSE WORKER DATA

👥

WORKERS SKILLSETS

COMPARE

WORKERS AVAILABILITY

✅

🔥

DISPLAY RECOMMENDATIONS TO THE SUPERVISOR

MORPHOLOGICAL CHART

concept 1 : cost effective

This concept is low on automation and hence cost effective would be fast to develop as it has less dependancies on technologies. It could be a good MVP solution test and iterate.

FUNCTION CARRIERS 👇 SOLUTIONS 👉	1	2
MONITOR MACHINES	Real time data received on machine performance	Based on Regular Inspection or worker observations
MACHINE AVALIBILITY	Read time data on machine availability and working condition	Supervisor maintained data on machines on ground
WORKER SKILLSET	Connected with HR qualifications database	Supervisor inputed data based on workers performance
WORKER AVALIBILITY	Data connected with shared workers calendars	Weekly Rota based availability
MATCHING RECOMMENDATIONS	Using ML / AI Algorithm	Supervisor manually matching based on observations and performance
DISPLAY	AR Head mounted display	Regular digital device display (TV, Wall Mounted Screen, Laptop, Hand Held Device)

MORPHOLOGICAL CHART

concept 2 : time efficient

This concept is fully automated and reduces the risk of data not captured correctly and time needed to do all the tasks manually. This uses Machine Learning and Artificial Intelligence and heavily depends on database. Thus would be more time taking but eventually the best option as that meets all the needs.

FUNCTION CARRIERS 👇 SOLUTIONS 👉	1	2
MONITOR MACHINES	Real time data received on machine performance	Based on Regular Inspection or worker observations
MACHINE AVALIBILITY	Read time data on machine availability and working condition	Supervisor maintained data on machines on ground
WORKER SKILLSET	Connected with HR qualifications database	Supervisor inputed data based on workers performance
WORKER AVALIBILITY	Data connected with shared workers calendars	Weekly Rota based availability
MATCHING RECOMMENDATIONS	Using ML / AI Algorithm	Supervisor manually matching based on observations and performance
DISPLAY	AR Head mounted display	Regular digital device display (TV, Wall Mounted Screen, Laptop, Hand Held Device)

CONCEPT EVALUATION

1 VS 2

Looking at the results it is clear that automation route is better than manual. Keeping in mind the irony of automation the suggestion is to be human-led technology-centred so the final decision is still on the supervisor but reduces the mental overload of managing huge amount of data on day to day basis.

CRITERIA	WEIGHTING (1 to 5)	CONCEPT 1		CONCEPT 2	
		VALUE (1 to 5)	WEIGHTED VALUE	VALUE (1 to 5)	WEIGHTED VALUE
Machine Information	5	4	20	4	20
Worker Information	5	4	20	4	20
Recommendations Accuracy	5	3	15	5	25
Return On Investment	3	4	12	3	09
Training Effort	2	5	10	2	04
Time Efficient	3	2	06	4	12
			83		90

AUTOMATION STRATEGY

Using the types and levels of automation framework, we can think about how to best use automation to further rather than hinder human performance

“

...a device or system that accomplishes (partially or fully) a function that was previously, or conceivably could be, carried out (partially or fully) by a human operator.
Parasuraman & Riley (1997)

”

LIST OF FUNCTIONS	SHOULD IT BE AUTOMATED	TYPE OF AUTOMATION	LEVEL OF AUTOMATION
MONITOR MACHINE PERFORMANCE	Yes, as this mainly needs analysing machine data	Acquistion automation	8 The system executes the checks automatically and then only informs if asked
IDENTIFY MACHINE FAILURE	Yes, depending on just worker & supervisor observations is not enough as it could delay the plant function	Analysis automation	7 The system makes the necessary checks automatically and notifies incase of machine failure.
ANALYSE MACHINE REQUIREMENTS vs AVAILABILITY	Yes, as it needs to not only identify the right machine and check its status but also confirm if they are available.	Decision automation	3 The computer narrows the selection of alternative machines down to a few based on the set criteria
ANALYSE WORKERS SKILLSET vs AVAILABILITY	Yes, it would be easy to connect to HR data instead of manually finding the right skills.	Analysis automation	10 The computer decides everything, acts autonomously, ignoring the human.
MATCHING RECOMMENDATIONS	Yes, it would be great to have a working recommendation engine where all the above data is analysed and using AI the right matches are suggested.	Decision automation	7 Using AI the system would match the right machine with the right set of workers automatically and then just notify the supervisor the recommendations.

PRIMARY EVALUATION

Primary evaluation criteria to identify the human performance consequences.

PRIMARY CRITERIA	DESCRIPTION
MENTAL WORKLOAD	With the help of automation most of the manual tasks like checking the machine performance, identifying issues and managing workforce have become really easy which helps in reducing mental workload.
SITUATION AWARENESS	Irrespective of the automation it is important for the supervisors to be situationally aware of the issues that might come up or to check if the working is going as per the plan and jump in if needed.
COMPLACENCY	With the range of automated tasks it is easy for the workforce to get complacent as the day to day manual tasks will reduce.
SKILL DEGRADATION	When the level of automation increases the skills are replaced with maintaining and ensuring that the work is running smoothly but it would also result in some extent of skill degradation.

SECONDARY EVALUATION

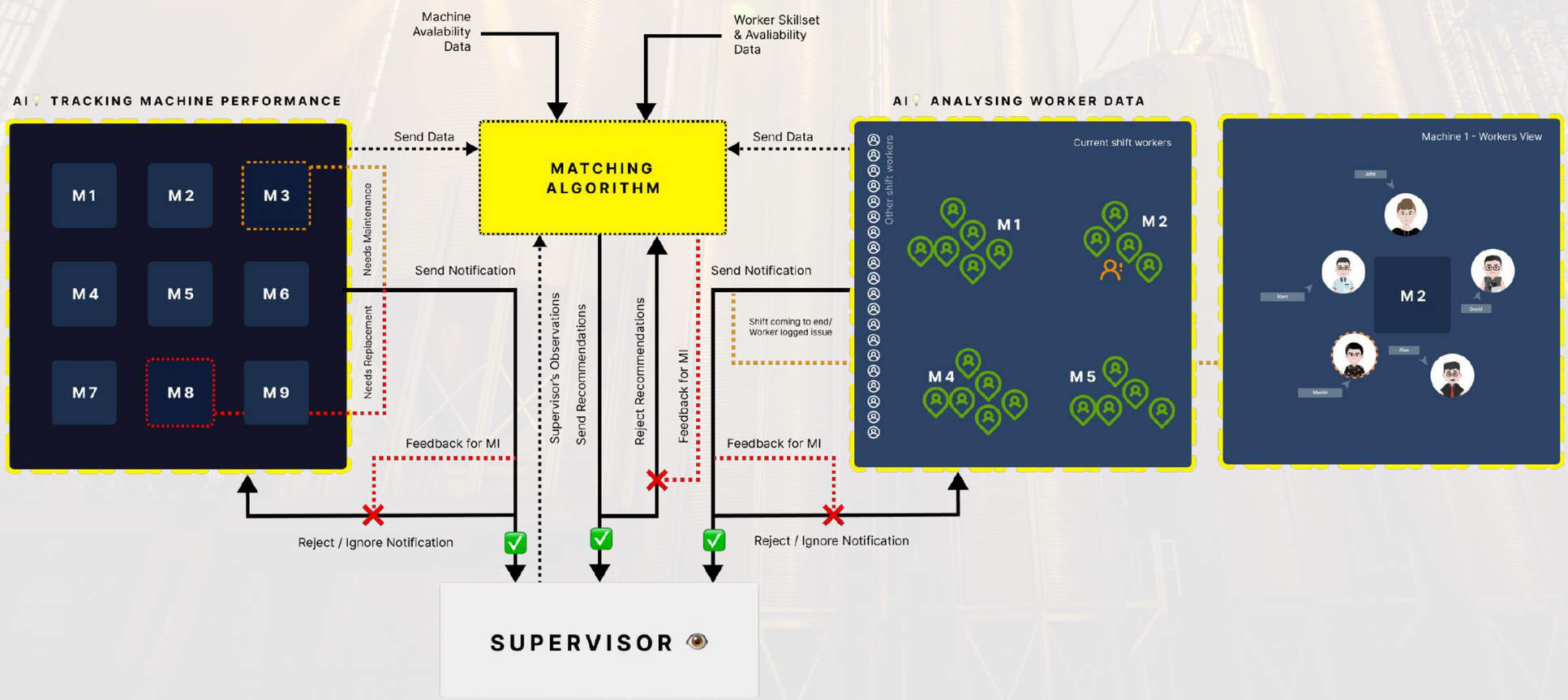
Secondary evaluation criteria to identify how reliable the automation is and what are the costs
for decisions and outcomes

SECONDARY CRITERIA	DESCRIPTION
AUTOMATION RELIABILITY	In the proposed design concept the automation range is high as it depends on AI model to provide the correct match and come up with recommendations based on different set of data points. Hence it is important that the system doesn't fail in between as it will have a chain effect. Potentially there could be an issue if first set of data is wrong and hence the matches don't align with the requirements. It is important that the workers / supervisors are not complacent only depending on the system generate recommendations and cross check if needed.
COST OF ACTION OUTCOMES	The cost of failing to detect issues in the data early on can result in potentially higher risk of wrong predictions.

INTERACTION STRATEGY

Principles of mixed-initiative interfaces used in the automation plan

Mixed initiative interfaces provide further guidance on how to build systems that couple principles of direct manipulation with automation.



INTERACTION STRATEGY PRINCIPLES

UTILITY

By having automated monitoring of machines and workers it reduces the time and efforts of the supervisor and possibilities of mistakes in reporting and is much more quicker than having to individually keep track of everything.

BALANCE

Would provide real-time machine information, failures, and incident reports to help the supervisor make informed decisions.

CONTROL

Head Mount Interface would be intuitive and easy to use, with features and functions that would meet the supervisors' needs. The user would be given the option to accept or reject suggested matches (machine + worker) and to provide feedback, which will help improve the accuracy of the matches over time.

UNCERTAINTY

When there is not enough data (system failure / connection issues) the AI would present the available info to the supervisor so that an informed decision could be taken. Ensuring uncertain data is not used to generate recommendations which could potentially lead to more issues.

INTERACTION STRATEGY CONSIDERATIONS

HUMAN-AI TEAMING ISSUES

It is important that the supervisor has trust on the matches suggested by AI. In order to achieve that level of teaming AI should generate more than 1 match and give reasoning behind the recommendations as that helps build trust.

ALIGNMENT BETWEEN USER AND AI

AI would automate data gathering and scanning making it easy for the moderator to access info to make a judgement. Having transparency in how the matches are generated would help build trust and by gathering feedback from the supervisor based on real time updates would help in machine learning which would in turn improve the matches in future.
Having this level of control ensures the supervisor and the AI have a common objective which is reducing downtime when machines don’t work.

MODEL UPDATES

With any new update it is important that there is training provided to the supervisors to explain and manage expectations on what has changed and why.

Adhering to these principles would ensure the task is carried out in a collaboration.

AI INTERPRETABILITY

the nature of the data the manufacturing plant processes is likely very **high dimensional** as multiple data will be processed at the same time

MACHINE DATA

Machines could stop working, slow down, need replacement, need maintenance. To track this level of information it would need to process individual functioning of these machines. Thus making the data very high-dimensional.

WORKER DATA

Workers could be off duty, change the shift, cancel shift, undergo training thus improving their skillsets. This information is very important for setting up rota schedule every week and also while assigning work to individuals. That means all this stream of information coming from different dimensions are equally important for everyday working.

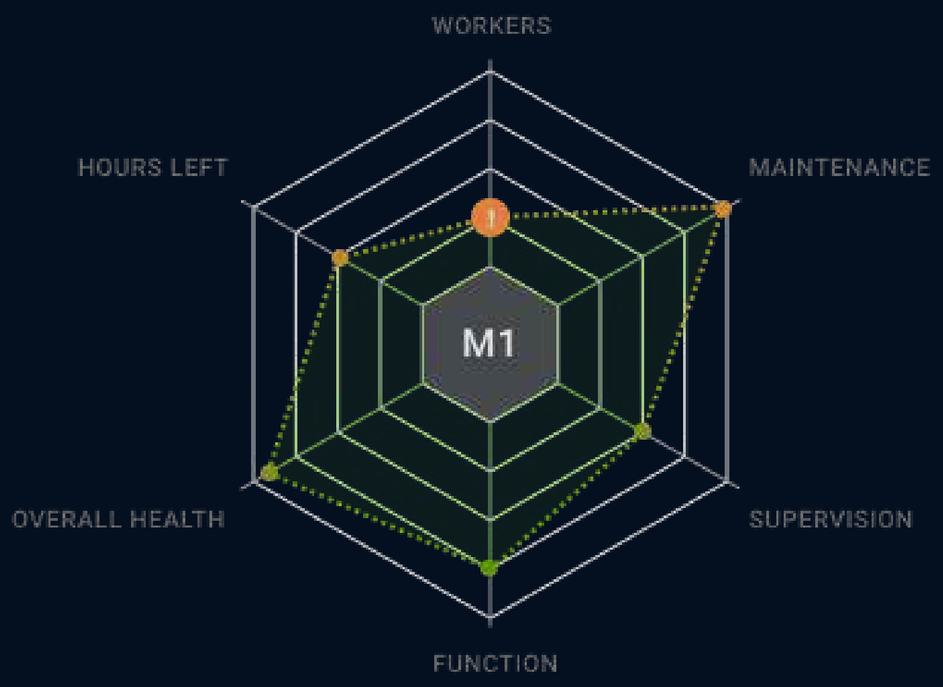
Both the machine data & worker data mentioned above could rapidly change based on individual conditions and hence it is highly uncertain.

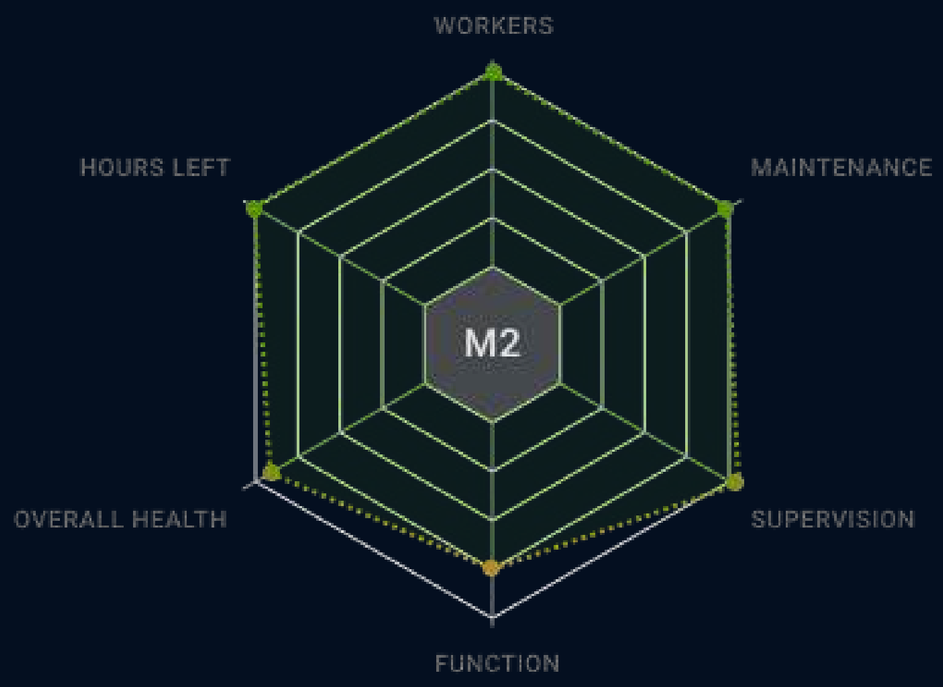
AI INTERPRETABILITY

level 1 : radar charts

I have used sample data to convey the idea. Using this overview the supervisor would be able to spot any important information quickly. In the above example it shows there is some important info on workers section which would be level 1 information.

Positive: Good for quick glance as it is easy to notice where the issue without getting lost when there are different layers of information.
Negative: The data on its own could be vague and not provide any insight unless the supervisor investigates **information regarding workers skill sets and availability**



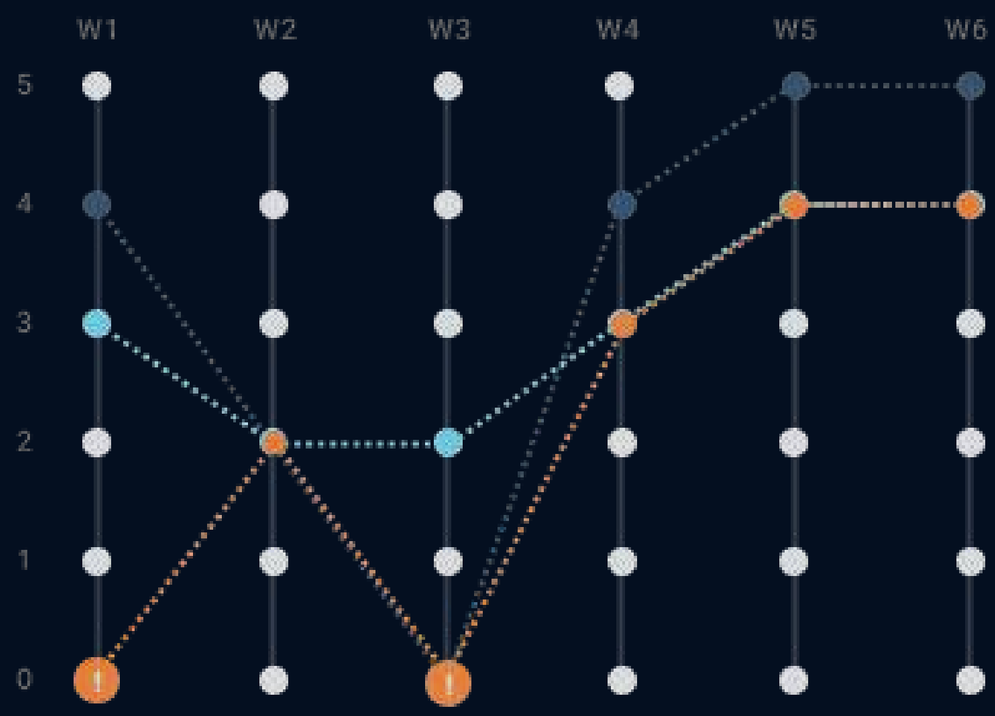


AI INTERPRETABILITY

level 2 : p**arallel coordinates**

I have used sample data to convey the idea. Using this visualisation the supervisor would be able to spot detailed information which requires his action. In the above example it shows that the shift of Worker 1 & Worker3 is coming to end and that in order to continue the working the supervisor can confirm the next set of workers are in line.

Positive: Good for deep dive and to compare multiple sets of data.
Negative: Multiple data points means there are more moving parts which would need some training for first time users.



AI INTERPRETABILITY CONSIDERATIONS

INTERPRETABILITY PROBLEM

A lot of data is being processed by the matching algorithm before making any recommendations which increases the risk if anything changes.

APPLICATION EVALUATION

Evaluate the interpretability with respect to how it solves an actual application with the help of a subject matter expert

METHOD TO ENABLE

Showing a decision tree or data suggesting why these recommendations were made would help build trust.

EVALUATION OF SOLUTION

A feedback loop after the matches are suggested will help machine learning and future recommendations

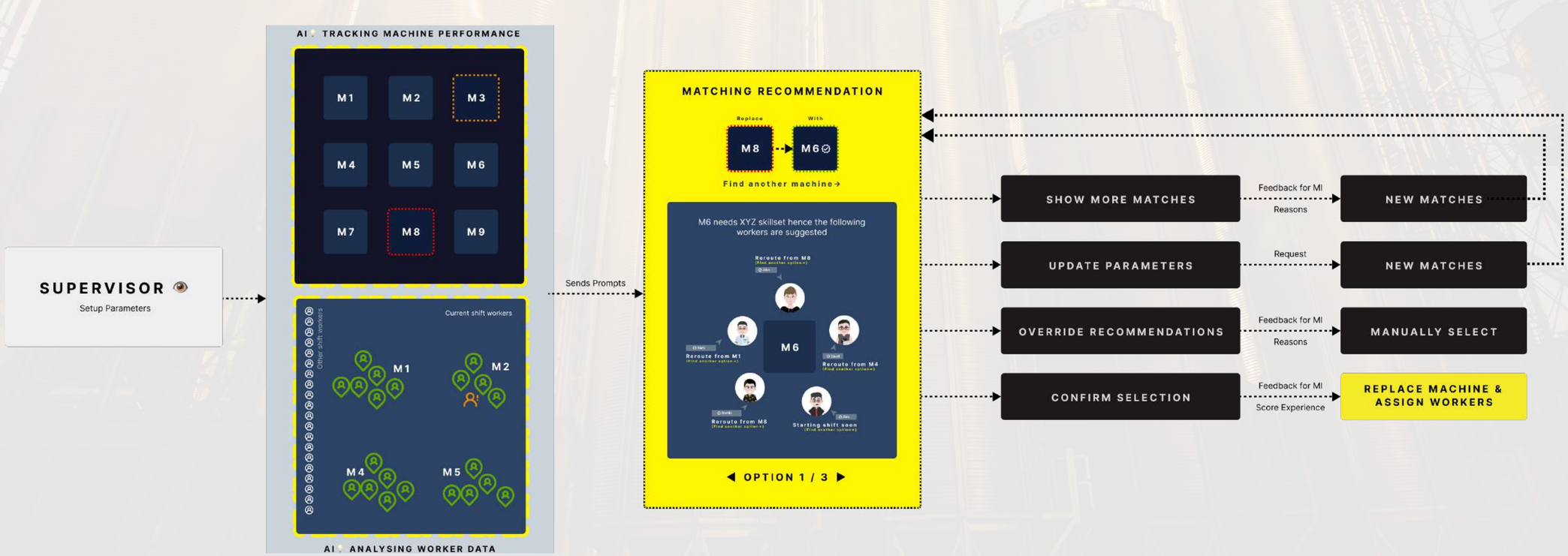
HUMAN EVALUATION

Another possibility is to study the efficacy of the interpretation of the AI through experimental investigations of more abstract tasks.

SHARING OF CONTROL AND USER AGENCY

Principles of mixed-initiative interfaces used in the automation plan

Methods for visualising high dimensional data and ways of enabling users to interpret an AI, both help users get a sense of agency and provides them with a means to control an AI system.



SHARING OF CONTROL AND USER AGENCY

SHARING OF CONTROL

In the proposed system, the control is **shared between AI and the supervisor**.
This is a **Loose rein control** as AI will do a bigger part by monitoring the machine performance to identify failures, analyse workers skillsets vs availability based on the data inputed by HR, rota schedule, supervisor notes & worker performance before recommending matches to the supervisor. This is an **operational type of control** as the supervisor is involved in setting up the parameters and making the final decision on matches and the AI is only providing data support by monitoring, analysing and suggesting the solutions.

DIFFERENT WAY OF SHARING

If the system is shifted to **strategic instead of operational** it would reduce the role of the supervisor further where the **system will automatically replace the machine and change the workers** based on the AI analysis.

IMPLICATIONS (If any)

By changing the shared control by making the system **more autonomous can have huge implications** where there are unexpected situations like (data not being updated / network issues) It could also impact and workers and supervisors could have trust issues with the recommendations which would be done based on **skewed data**.

SHARING OF CONTROL AND USER AGENCY

SENSE OF OWNERSHIP

As the **data is very high-dimensional**, the supervisor will have a sense of control if they **have control over the final decision** instead of the machine directly taking action based on the data alone.

PROMPTS

With automation, it is great to have regular prompts of what is happening in the background so the **supervisor is aligned** with what is happening and can take actions early on if needed.

This also **helps building trust** when the suggestions are provided by AI as the supervisor is **in the loop throughout**.

CONTROL AND USER AGENCY CONSIDERATIONS

RISK OF INACCURACY AND THE USE OF MACHINE LEARNING

There is certainly a risk of inaccuracy where the recommendations are based on high-dimensional data hence having automation administered by the supervisor who can take the final decision will reduce the risk. It is important that the machine learning / machine teaching is implemented in this kind of automation as each result could have multiple actions such as:

REQUEST MORE MATCHES / MATCH OPTIONS – In case the supervisor is unsure about the results or needs more options before they take the decision. This can be captured as confidence scores and the system will start learning which was the better match and update for future recommendations.

UPDATE PARAMETERS – Catering to the changes it is necessary to have an option to update the set parameters so the matches are not affected. This could also be triggered if the matches are not as expected and can help ML for future.

OVERRIDE RECOMMENDATIONS – The supervisor has to make a different match based on some unexpected situation. It is important to gather feedback after these actions so it can be used for future matches.

CONFIRM SELECTION – In a positive scenario where the AI matches are confirmed by the supervisor having a rating on trust and experience can help ML for future.

SYSTEM BOUNDARY AND RISK ANALYSIS

To prevent unanticipated system outcomes, we can use methods from design engineering and human factors to map out systems and risk assessment, allowing us to reason about acceptable risk levels and continuously measure the level of risk in a human AI system.

PART OF THE SYSTEM

Workers
Supervisors
Machines
Tech like AI / ML
Power Supply
Network
Manufacturing Plant
Worker Availability
Worker Skillset
Machine Health
Health & Safety Regulations
Any material, parts of the machine needed for the factory

NOT PART OF THE SYSTEM

Any other Regulations that might impact but are not very common to change.
Third party agencies
Workers family
Anything happening outside of the manufacturing plant

SYSTEM BOUNDARY AND RISK ANALYSIS

Select one system mapping method and use it to map out the automation function
I have selected process diagram to map the system as there are multiple actors and factors that could impact the process.

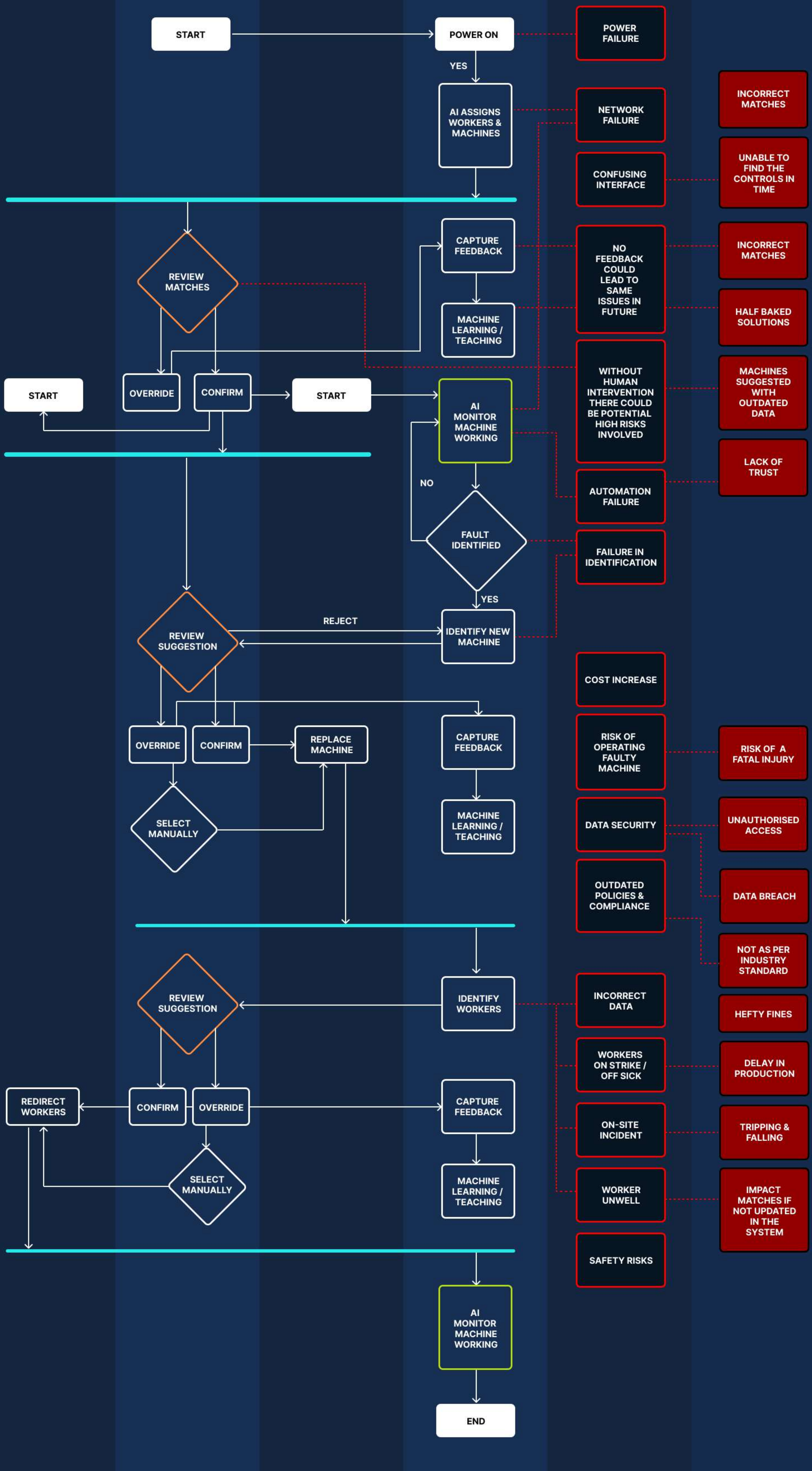
Was your choice of system mapping technique useful for assessing risks later? If so, why? If not, why not?
Swim lane process diagram helped me deconstruct the entire process to be able to figure out if anything is missed or if there were any risks associated with any steps.

Is your system boundary truly encompassing all relevant factors, including policies and regulation?
I have kept the regular policies & regulation within the boundary which have direct impact but avoided any external factors that could potentially change.

What hazards did you identify? Were there hazards that were not exposed?
I initially mapped out all the potential hazards and risks in the process diagram as that way I could quickly analyse risks associated in all these steps.

Determine an acceptable level of risk and use one risk assessment method from this module to assess the five key risks you foresee. Was your choice of risk assessment method useful for assessing the risks in your system? If so, why? If not, why not?
I have used the SWIFT analysis as it was helpful to identify the necessary steps that needs to be taken before hand to avoid system failure.

WORKERS	SUPERVISOR	MACHINES	AUTOMATION	POTENTIAL RISKS	POTENTIAL HAZARDS



IDENTIFIER	WHAT IF SCENARIO	RISKS & HAZARDS	RELEVANT CONTROLS	ACTIONS	RANKING
1	WHAT IF... the AI doesn't identify machine failure?	Workers could keep working in a potentially risky situation where they could be exposed to hazards. **DELAY IN PRODUCTION** **COST INCREASE** **RISK OF A FATAL INJURY**	Physical checks on floor to ensure the machines are working as expected.	Train supervisors & workers to spot any failures as a contingency plan. Make sure to update the machine data regularly to avoid any slips.	HIGH - 5
2	WHAT IF... the worker data is incorrect?	Workers could have earned new skills which expand their knowledge about new machines or are injured / late / off sick which could all potentially affects the selection process **INCORRECT MATCHES** **LACK OF TRUST** **RISK OF A FATAL INJURY**	Connect the HR database so that it automatically updates in the system	Add mandatory supervisor approval before any changes to ensure the suggestions are correct. Train supervisors so that they can spot if there are any issues with the recommendations and override if needed.	HIGH - 5
3	WHAT IF... the AI recommendations fails to improve over time?	**LACK OF TRUST** **DELAY IN PRODUCTION** **SAFETY RISKS**	Way to capture feedback at all points in the journey	Add an incremental machine learning loop so that over the time the suggestions would improve. Collect feedback regularly and feed the data back in the system.	LOW - 1
4	WHAT IF... the AI fails to consider latest policy & compliance changes?	**NOT AS PER INDUSTRY STANDARD** **HEFTY FINES** **DELAY IN PRODUCTION** **SAFETY RISKS**	Way to update the compliance & policy changes regularly	Regularly check for updates and train supervisors to update the latest policy and compliance guidelines into the system.	MEDIUM - 3
5	WHAT IF... the AI network fails and system doesn't connect?	**AUTOMATION FAILURE** **FAILURE IN IDENTIFICATION** **DELAY IN PRODUCTION**	Regular monitoring of system and network to avoid failures	Train supervisors & workers to work out a contingency plan if the system is not back up in few hours to avoid production delays.	LOW - 1

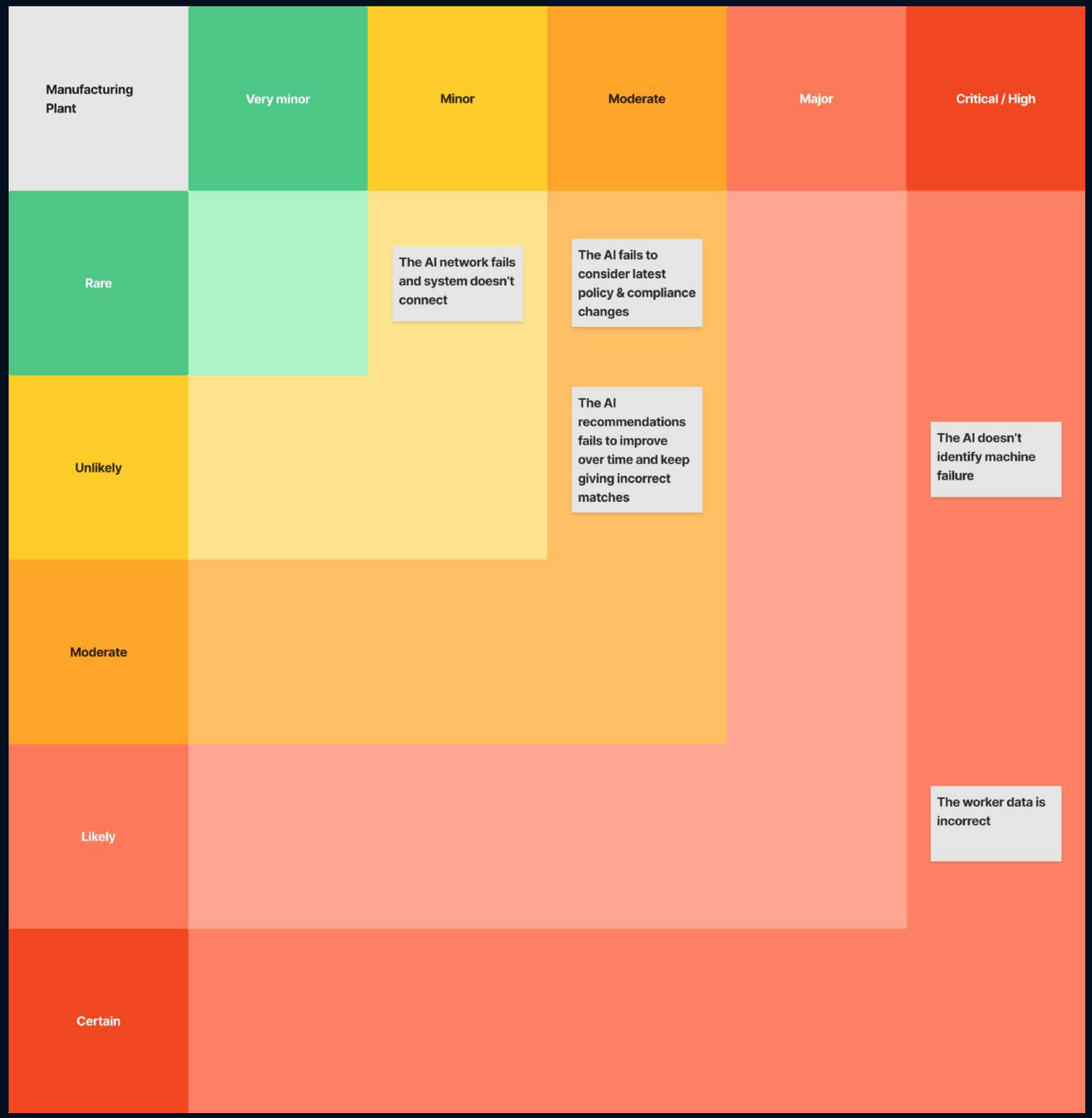
SYSTEM BOUNDARY AND RISK ANALYSIS

How would you redesign your automation function to reduce the level of risk?

After completing the risk analysis it was clear that certain risks can be mitigated by adding some rules and by ensuring to run some protocols.

By making the changes like - adding a feedback loop to improve the quality of recommendations, to creating a contingency plans so in case of any failure the product won't be impacted.

It was also clear that having human control and training the supervisors / workers is really important as on the floor decisions would be needed in case of any issue due to automation.



VERIFICATION CROSS REFERENCE MATRIX

Finally, we can verify that requirements have been met and validate whether design, indeed, fulfils users needs and wants.

NO	REQUIREMENT	VERIFICATION	ALLOCATION	SUCCESS CRITERIA
1	AR head mount device must indicate that it is powered on	Inspection	User Interface	Visible power on status to avoid any confusion
2	Device must alert the supervisor if disconnected from the network	Demonstration	Wireless connectivity	Connectivity failure results in device alerting the supervisor so the right actions are taken quickly as that could impact the working in the plant
3	The device must estimate machine failure with a 95% true positive rate at a maximum false positive rate of 5%	Analyse	Machine Monitoring	Requirements met through analysing data for the decided time frame to get the average
4	The matches suggested by AI must be accurate based on latest info and updates	Test	Matching Algorithm	Consistent ability (>99%) to generate accurate matches as inaccurate matches can lead to trust issues and can increase the machine down time resulting in production delays

DEPLOYMENT STUDIES

of the verified and validated system

Since the tool will be released to all the supervisors in the manufacturing plant there will be opportunities for studying the system to further iterate and improve based on the research and feedback.

UNIVERSITY OF CAMBRIDGE

THANK YOU

AISHWARYA KARVE I CAMBRIDGE UNIVERSITY I FINAL REPORT

OVERVIEW PROBLEM FUNCTION AUTOMATION INTERACTION INTERPRETABILITY CONTROL RISK VERIFICATION