These projects were funded by the European Commission

Collaborative Intelligence for Safety Critical Systems: challenges and opportunities

CISC
Collaborative Intelligence for Safety Critical systems

H2020-MSCA-ITN-2020. No. 955901 &

Maria Chiara Leva

Technological University Dublin, City Campus, Ireland

H2020-Grant Agreement Number 957402

The people making this possible

Houda BRIWA

Hosting Institution: TU Dublin - Supervisor: Prof. Maria Chiara LEVA

Devesh JAWLA

Hosting Institution: TU Dublin - Supervisor: Prof. John D. KELLEHER

Carlos ALBARRÁN MORRILLO

Hosting Institution: POLITO - Supervisor: Prof. Micaela DEMICHELA

Aayush JAIN

Hosting Institution: IMR - Supervisor: Prof. Philip LONG

Joseph MIETKIEWICZ

Hosting Institution: HUGIN EXPERT A/S - Supervisor: Prof. Anders L. MADSEN

Chidera WINIFRED AMAZU

Hosting Institution: POLITO - Supervisor: Prof. Micaela DEMICHELA

Shakra MEHAK

Hosting Institution: Pilz - Supervisors: Prof. Michael GUILFOYLE & Jack COLLINS

Naira LÓPEZ CAÑELLAS

Hosting Institution: European DIGITAL SME Alliance - Supervisor: Prof. Aphra KERR

Ammar ABBAS

Hosting Institution: SCCH - Supervisor: Prof. John D. KELLEHER

Miloš PUŠICA

Hosting Institution: mBrainTrain - Supervisor: Prof. Ivan GLIGORIJEVIC

Doaa ALMHAITHAWI

Hosting Institution: MATHEMA - Supervisor: Prof. Stefano CUOMO

Inês FERNANDES RAMOS

Hosting Institution: UNIMI - Supervisor: Prof. Gabriele GIANINI

Collaborative intelligence

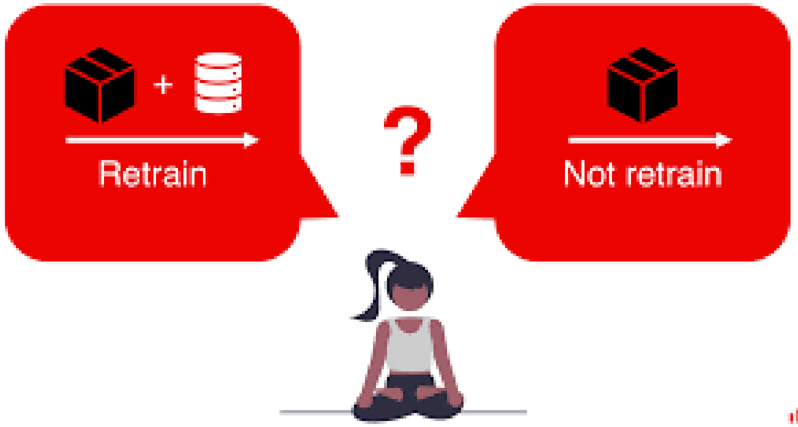
*"Organizations that use machines merely to displace workers through automation will miss the full potential of AI…Tomorrow's leaders will instead be those that embrace collaborative intelligence, transforming their operations, their industries and –no less important-their workforces."**

*A "human-centric" approach to AI that collaborate with humans rather than replace them.***

Human contribution:

Train

Explain



Sustain

* Daugherty, P.R.&Wilson, H.J., 2018. Human+Machine: Reimagining Work in the Age of AI. Harvard Business Press.

* * Leva, M.C., Podofilini, L. "Assessing Human Performance and Human Reliability in Collaborative Intelligence Scenarios: Upcoming Challenges and Opportunities" in Proceedings of ESREL2020-PSAM15

Collaborative intelligence

A “human-centric” approach to AI that collaborate with humans rather than replace them.

AI Systems’ contribution:

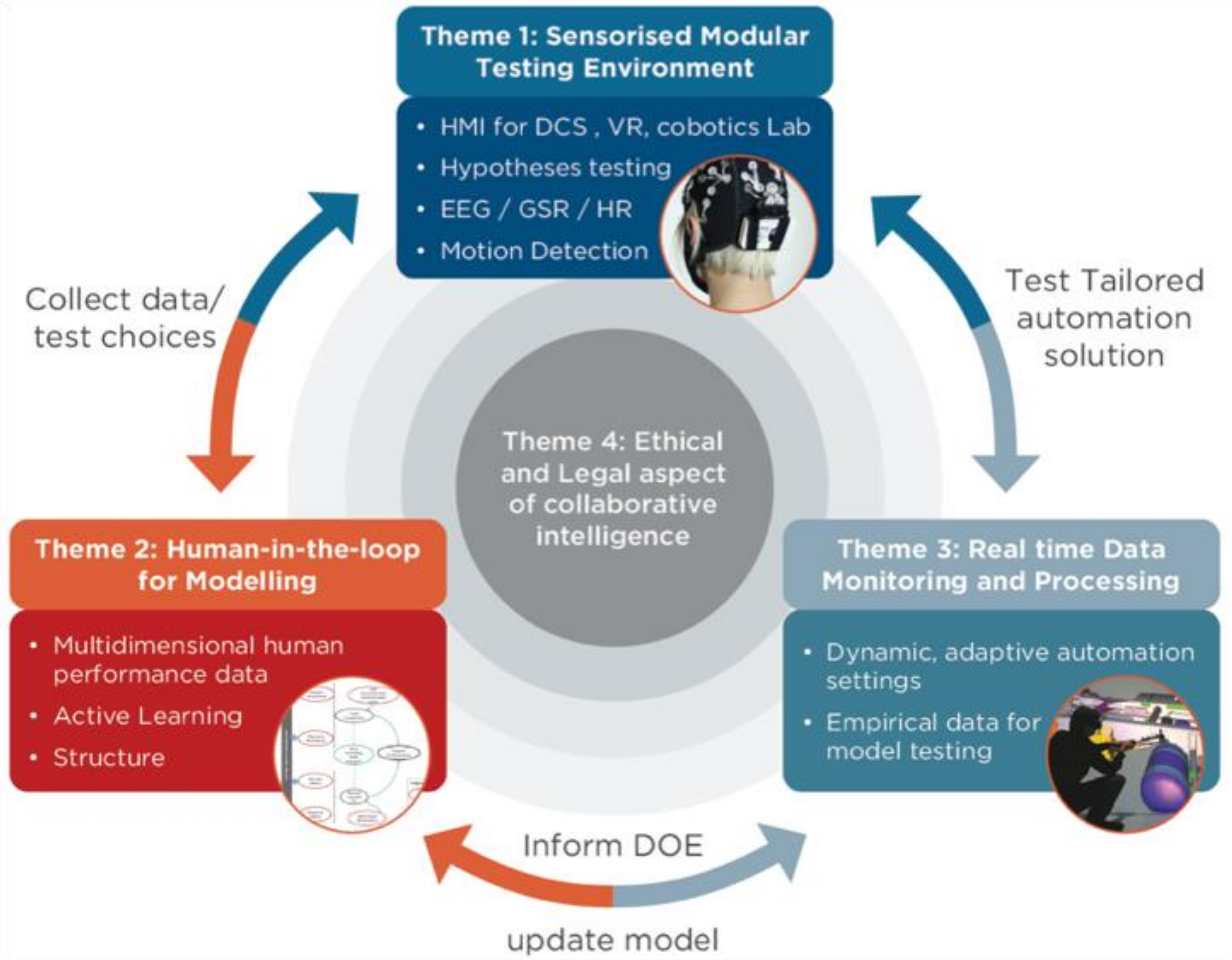
Amplify

Interact

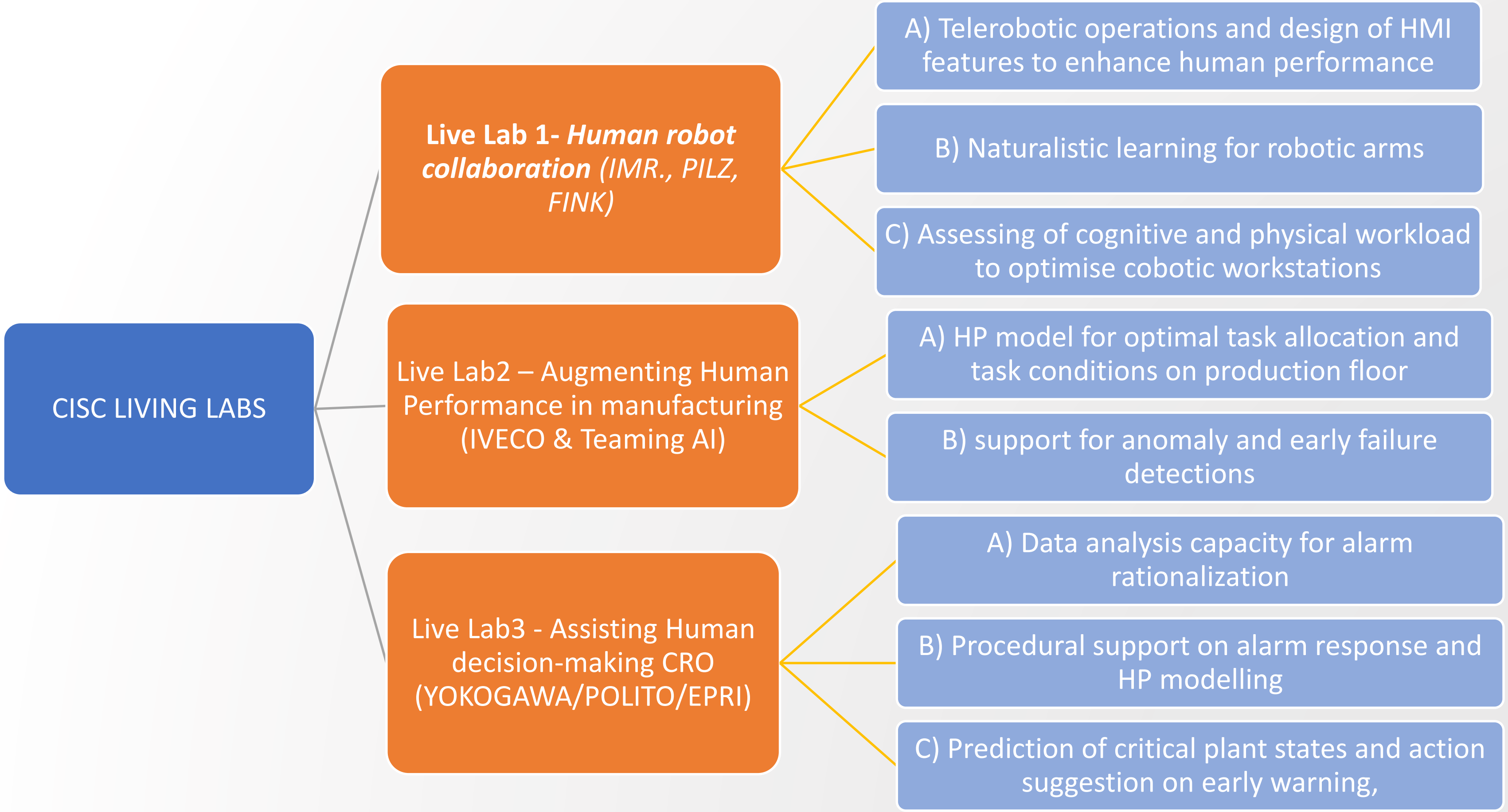
Embody

* Leva, M.C., Podofilini, L. “Assessing Human Performance and Human Reliability in Collaborative Intelligence Scenarios: Upcoming Challenges and Opportunities” in Proceedings of ESREL2020-PSAM15

The CISC Approach to DOE



The CISC Living Labs: collaborative intelligence examples



*Grant Agreement 955901- Collaborative Intelligence for safety Critical Systems (CISC). H2020-MSCA-ITN-2020.

CISC

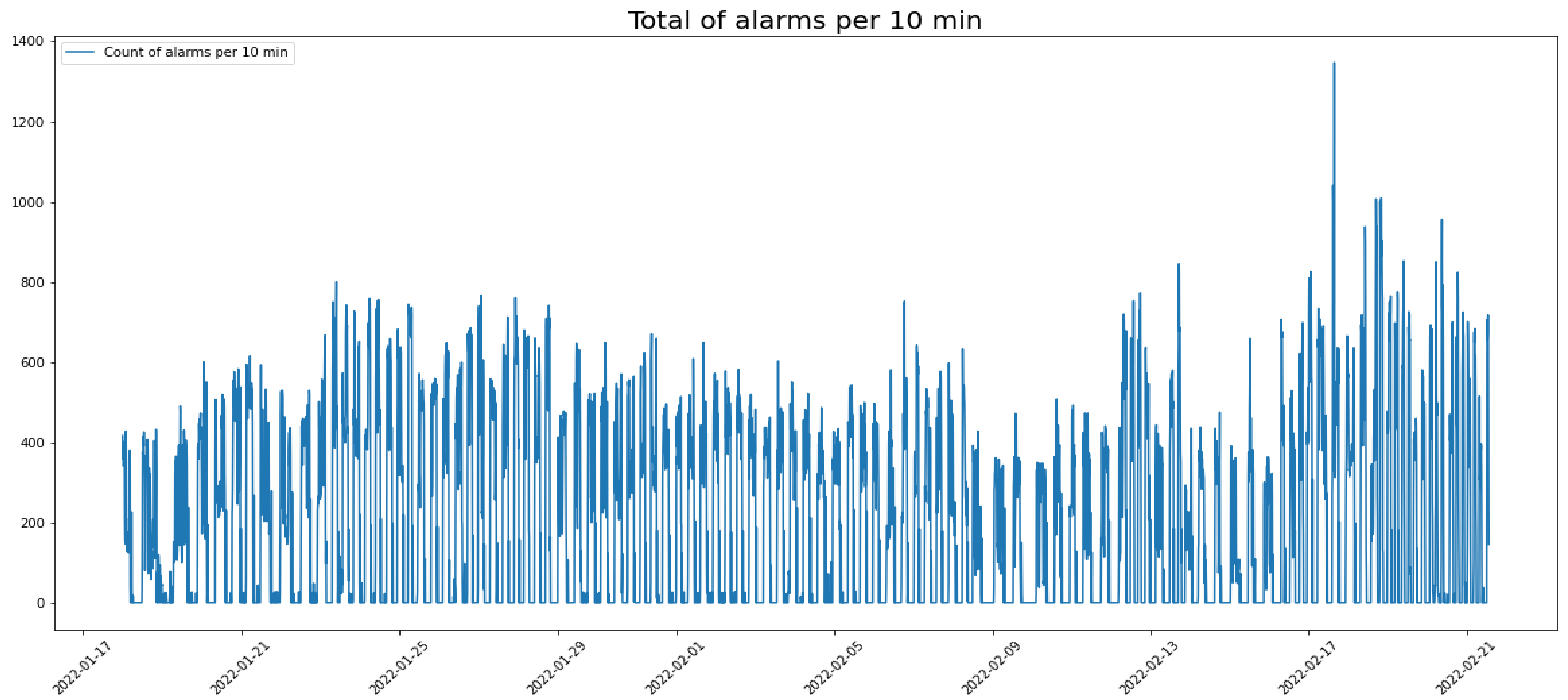
Collaborative Intelligence for Safety Critical systems

Collaborative intelligence in control room scenarios.

The use of a digital twin for testing different collaborative intelligence configurations

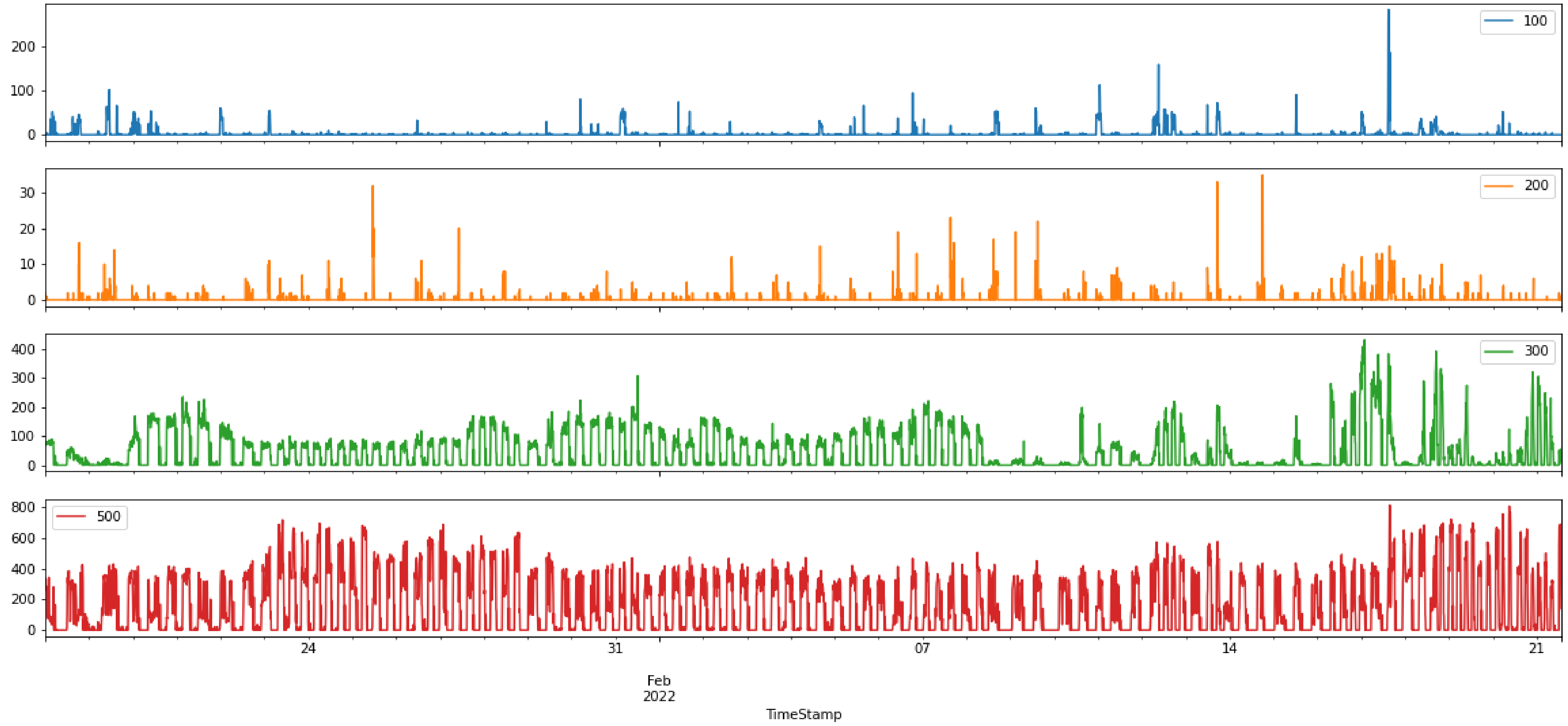
Live lab 3

Count of Alarms per 10 min



The graph illustrates the rate of alarms per 10 min over the period from 17 January to 21 February in a UK based Oil and Gas facility.

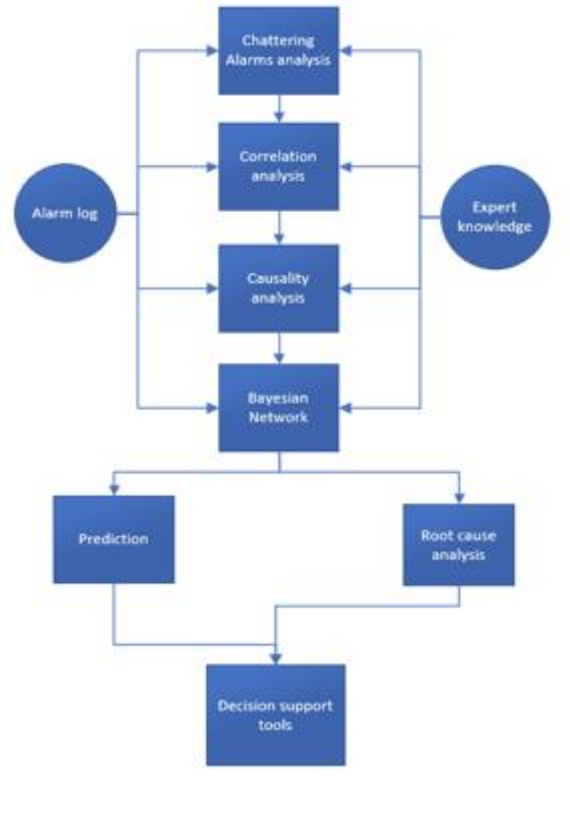
Count of Alarms per 10 min by Severity



The graph illustrates the count of alarms per 10 min over the period from 17 January to 21 February by level of Severity (100, 200, 300, 500)

Collaborative Intelligence for Safety Critical systems

Clustering and Bayesian Network proof of concept



1. Clustering of alarms based on correlation and identification of alarms related to the shutdown of the valves of the wellhead

2. Prediction of the trip of the shutdown valves of the wellhead 15 second before it happened

3. Root cause analysis to prevent the shutting down of the wellhead valves and remove redundant alarms

ciscproject.eu

CISC
Collaborative Intelligence for Safety Critical systems

1. Clustering of alarms

Identification of **clusters of alarms** to identify **scenarios**.

Possibility of **Grouping the alarms** base on high correlation to **reduce number of alarms** shown to the operator

These **groups of alarms** can be linked to a known cause and **labeled** using **expert knowledge**. The model can then display the causes of the alarms and assist in decision making in case of cognitive overload.

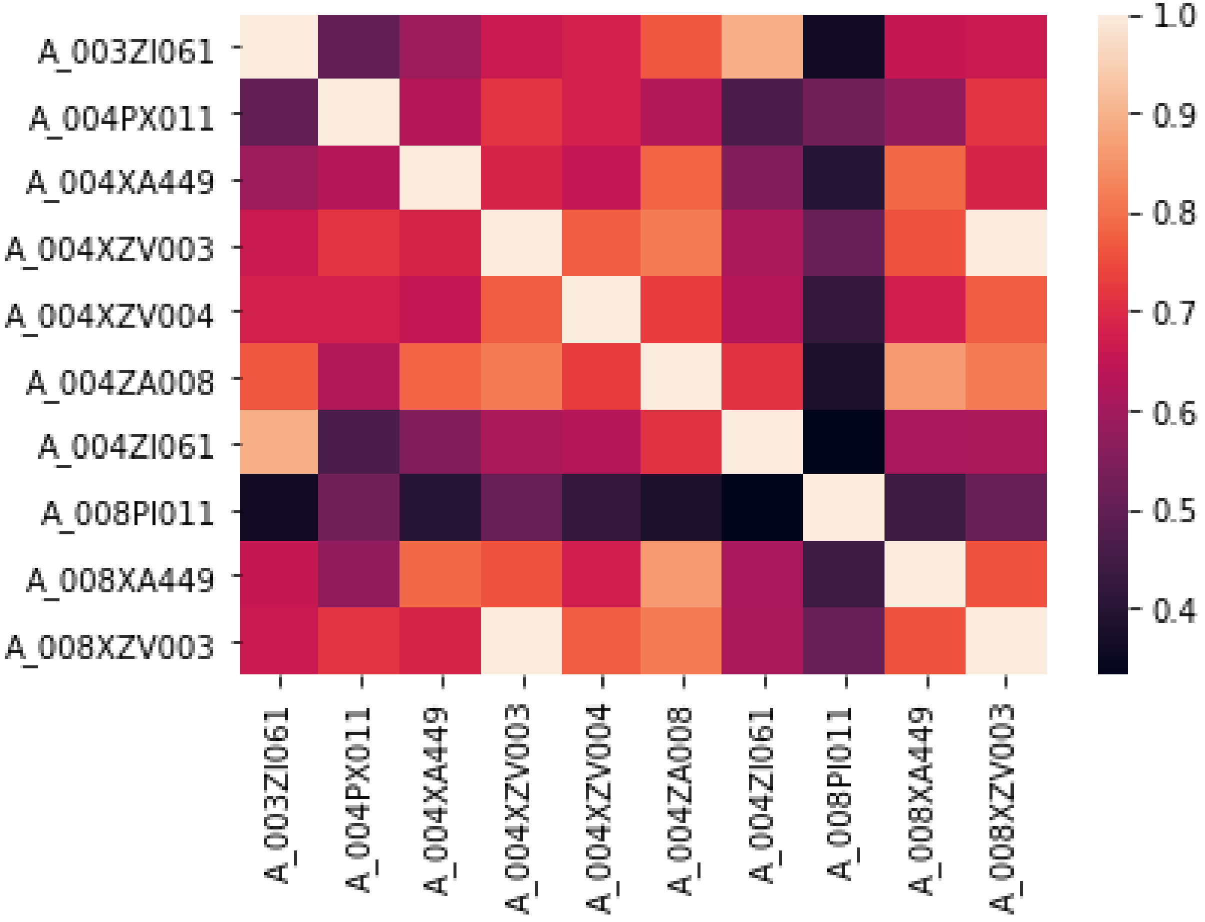
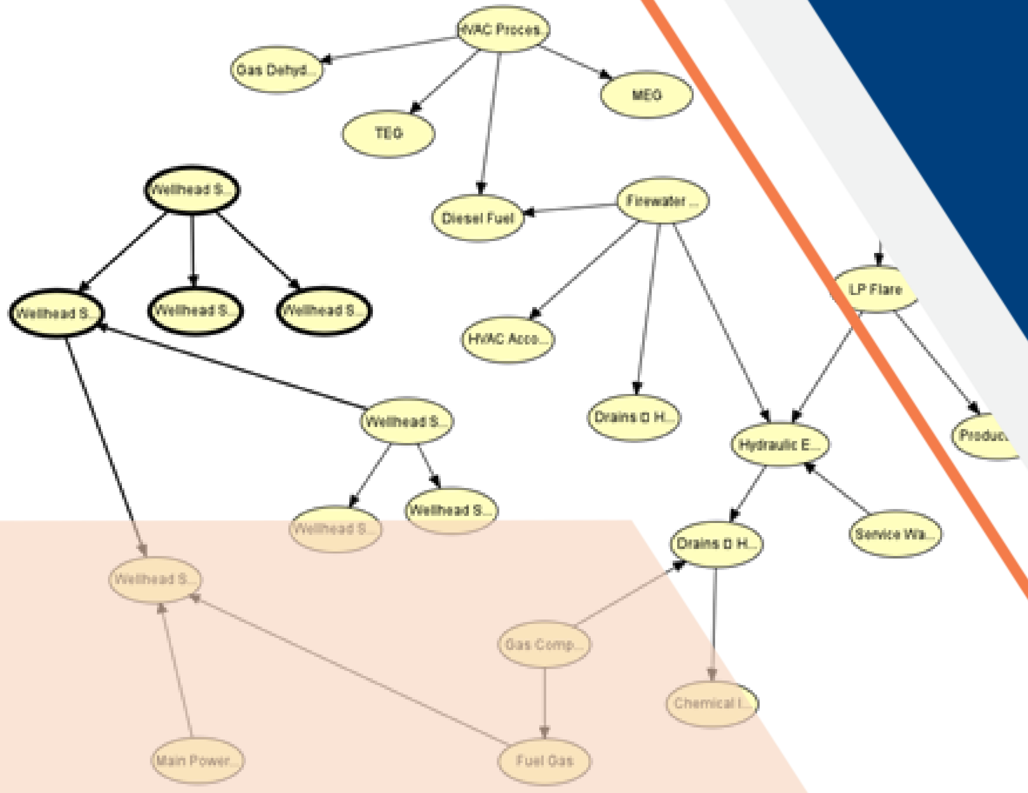


Figure: Correlation map between alarms of the wellheads. Lighter colors denotes higher correlations

Collaborative Intelligence for Safety Critical systems

Bayesian Network



Powerful machine learning technique to model causal interaction between variables

BN model interaction between alarms in the systems. Can be used to predict alarms, Trips and identify root causes. Can Estimate risk and cost of a process upset.

The model allows Transparency in reasoning and trustworthy decision

Can be use for a short cuts to becoming an experience operator thank to the decision-making models.

CISC
Collaborative Intelligence for Safety Critical systems

ciscproject.eu

3.Possible Root cause analysis

Redundancy between the alarms to predict the trip of the wellhead.

3 alarms to predict the trip with 90% chance **over 38**

Possibility to **reduce the number of alarm** display to the operator.

Alarm prioritization in terms of increase of probability of TRIP

Top 3 Alarms (026 Gas compression system)
A_026BPZI070 Low Low Gas compression pressure indicator
026BFI064 Open Alarm Gas compression press flow indicator
026BPI047 Hight Gas compression pressure indicator

Table: Alarm order in term of increase of probability of the TRIP of the wellhead.

Experiment: alarm management & intervention simulator

CISC

Collaborative Intelligence for Safety Critical systems

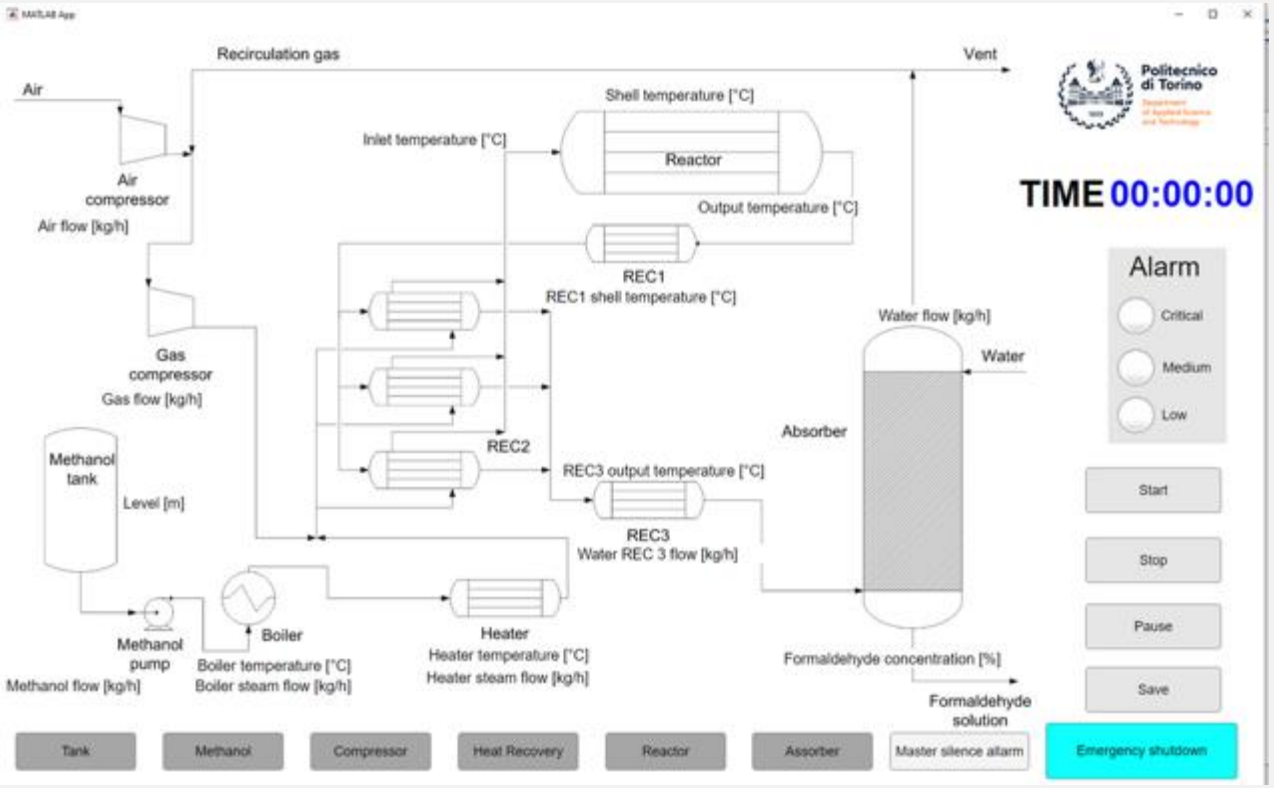
Live Lab 3 : Decision making support in control room

Factors varied: Alarm design

(Investigating problem 1)

Ack	Sound	Priority	State	Time	Tag	Section
☐	☐	3	Active	00:01:37	FAL11	Heat_Recovery
☐	☐	1	Active	00:04:38	TAH17	Assorber
☐	☐	2	Active	00:05:41	CAL20	Assorber
☐	☐	3	Active	00:05:53	FAL20	Assorber
☐	☐	1	Active	00:06:02	FAL16	Assorber
☐	☐	1	Active	00:06:26	TAH18	Assorber
☐	☐	2	Active	00:06:34	TAH20	Assorber
☐	☐	2	Active	00:06:54	TAH12	Heat_Recovery
☐	☐	3	Active	00:08:56	TAL15	Reactor
☐	☐	3	Active	00:09:11	TAH09	Heat_Recovery
☐	☐	1	Active	00:10:59	WAH07	compressor
☐	☐	3	Active	00:11:07	PAL15	Reactor
☐	☐	1	Active	00:15:48	TAL13	Reactor
☐	☐	1	Active	00:15:52	TALL14	Reactor
☐	☐	3	Active	00:15:57	PAH08	Heat_Recovery

Sound	Priority	State	Time	Tag	S
☐	1	Active	00:01:37	FAL11	H
☐	1	Active	00:04:38	TAH17	A
☐	1	Active	00:05:40	FAH08	H
☐	1	Active	00:05:41	CAL20	A
☐	1	Active	00:05:53	FAL20	A
☐	1	Active	00:06:02	FAL16	A
☐	1	Active	00:06:06	PAL08	H
☐	1	Active	00:06:26	TAH18	A
☐	1	Active	00:06:34	TAH20	A
☐	1	Active	00:06:54	TAH12	H



Type of plant - Chemical Process Industry

(Formaldehyde production).

Alarm flood condition: present

Micaela Demichela, Gabriele Baldissone, and Gianfranco Camuncoli.
Risk-Based Decision Making for the Management of Change in Process Plants: Benefits of Integrating Probabilistic and Phenomenological Analysis. Industrial Engineering Chemistry Research 2017 56 (50), 14873-14887

This experiment is to

CISC

- investigate the **impact of decision support systems** on control room operators in safety critical status,
- analyse the **different factors** impacting ***its ability to perceive and then respond*** (conduct actions on the monitor) to ***critical alarms***.
- There are four groups of participants (with different level of HMI support) and 3 scenarios with different level of complexity

01
Paper procedures

02
Alarm rationalization + Paper procedures

03
Alarm rationalization + Screen based procedures

04
Alarm rationalization + Screen based procedures + AI recommendation support

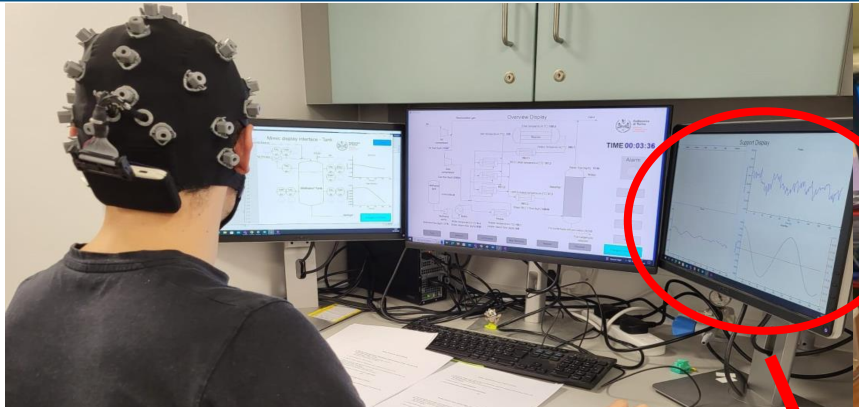
What does human in the loop mean?

- There are different types of “humans” in the machine learning loop
- human-in-the-loop decision-making is where content is flagged by the AI and human moderators review what has been flagged and confirm whether the machine was correct in order to enhance the algorithm's decision-making? (**this is one of the most widely used concept..but often not working well..)**

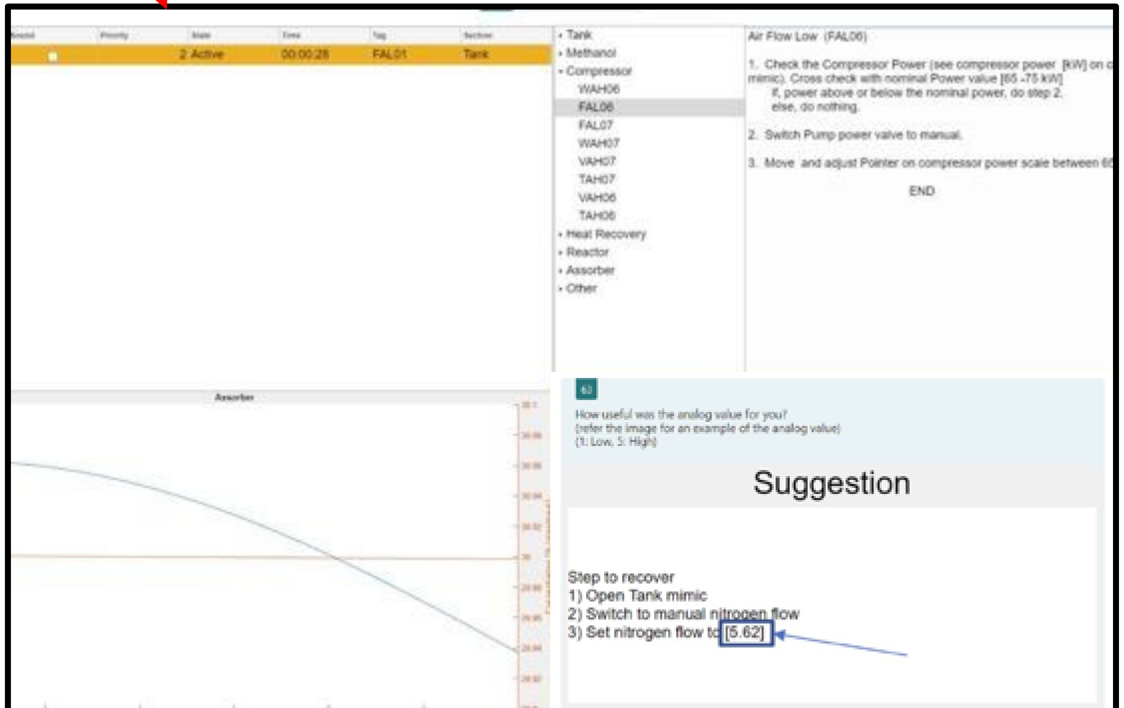
True HITL automation allows human intervention to execute actions and control the entire workflow. By allowing ad hoc application of human judgment, it’s more flexible and powerful. (Forbes technology council 2022)

The support Interface:

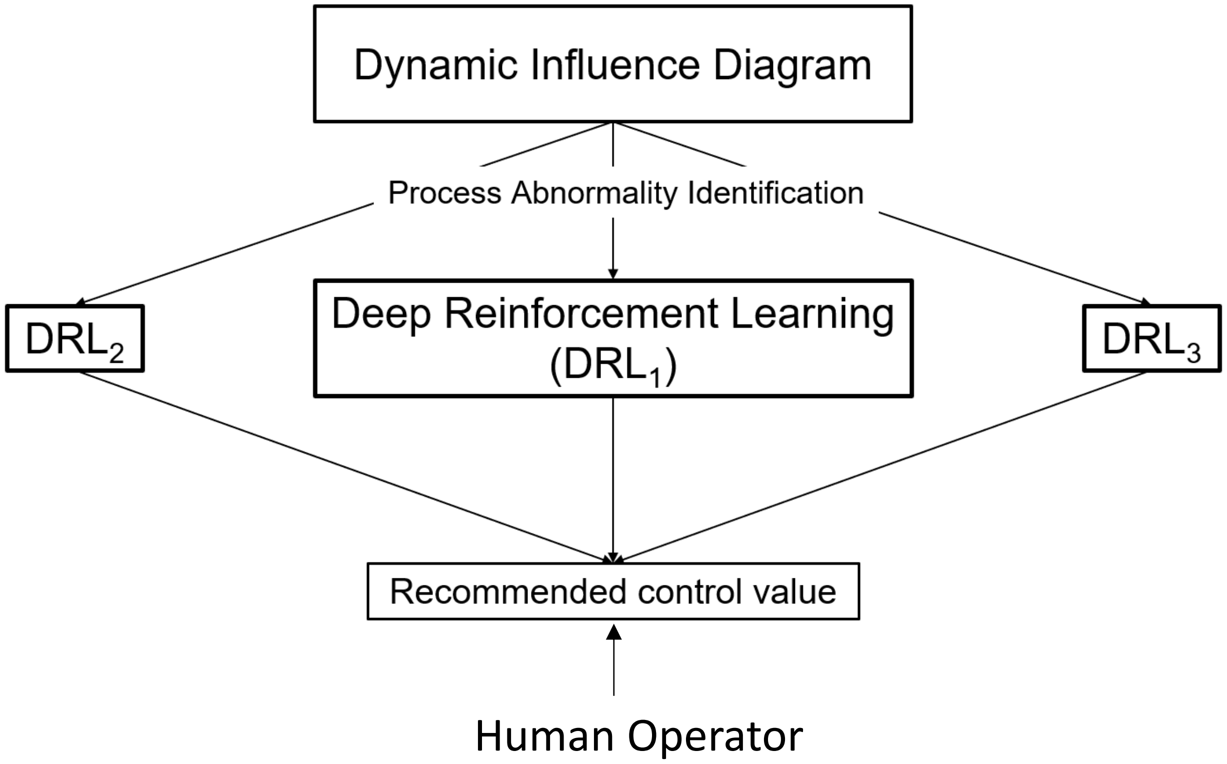
For GROUP 4 only it contains an AI generated recommendation system

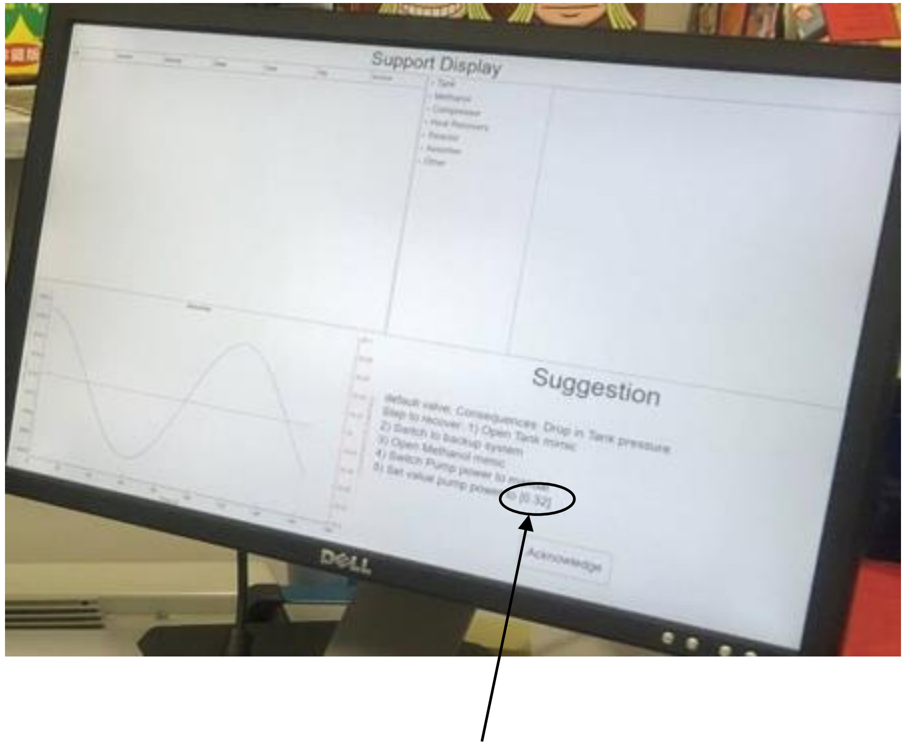


The support interface will appear on the right monitor, and it shows 4 sections.
The top left, shows the list of alarms and their different characteristics (name, state, priority, time, tag, section and acknowledgement case where the participant should click to acknowledge it).
The top right, shows the procedures section. the participant should click on the specific section then the specific alarm to view its corresponding procedure.
The bottom left, is a graph that shows and the flow of water and product concentration in the absorber.
The bottom right, is the AI recommendation system

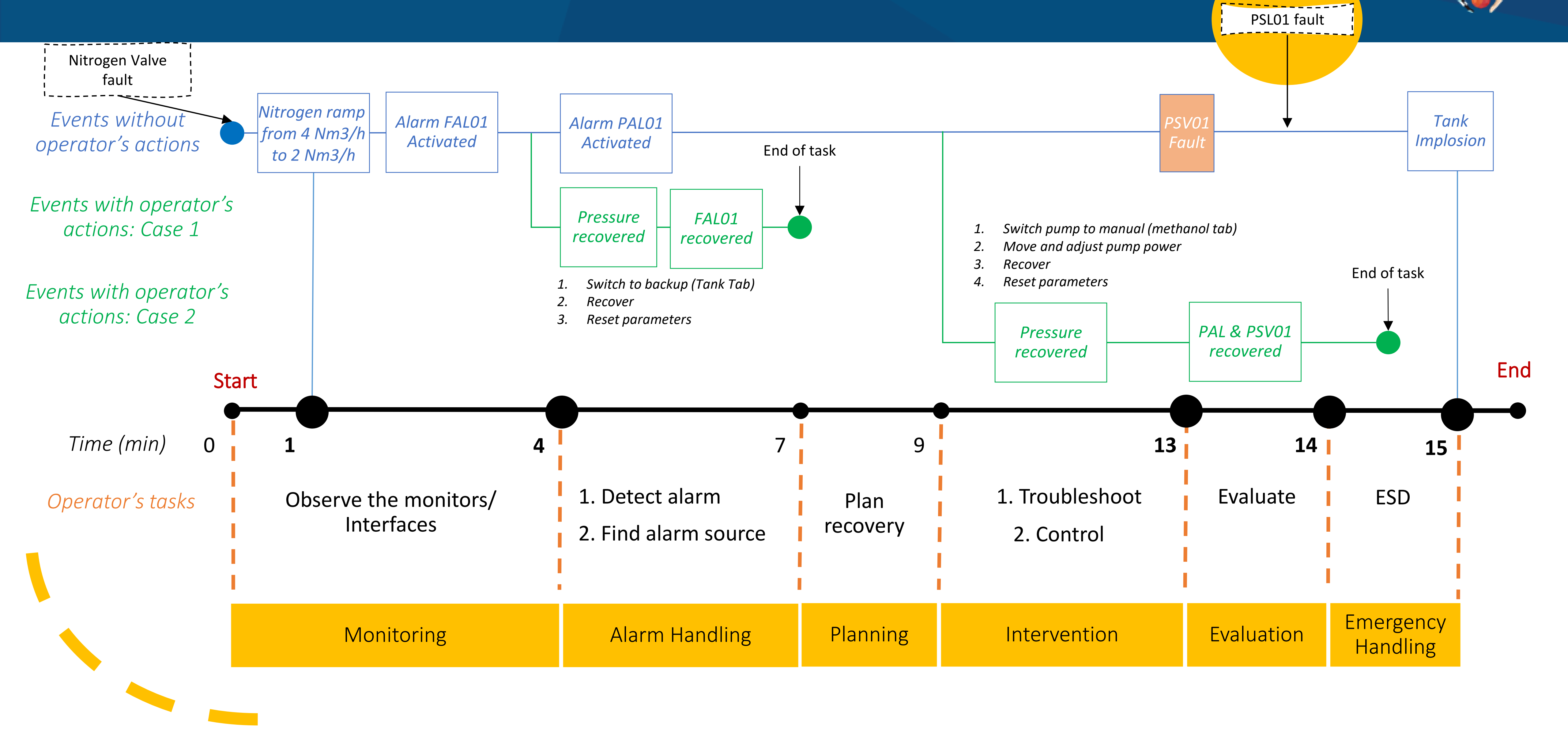


AI-Enhanced Recommendation System

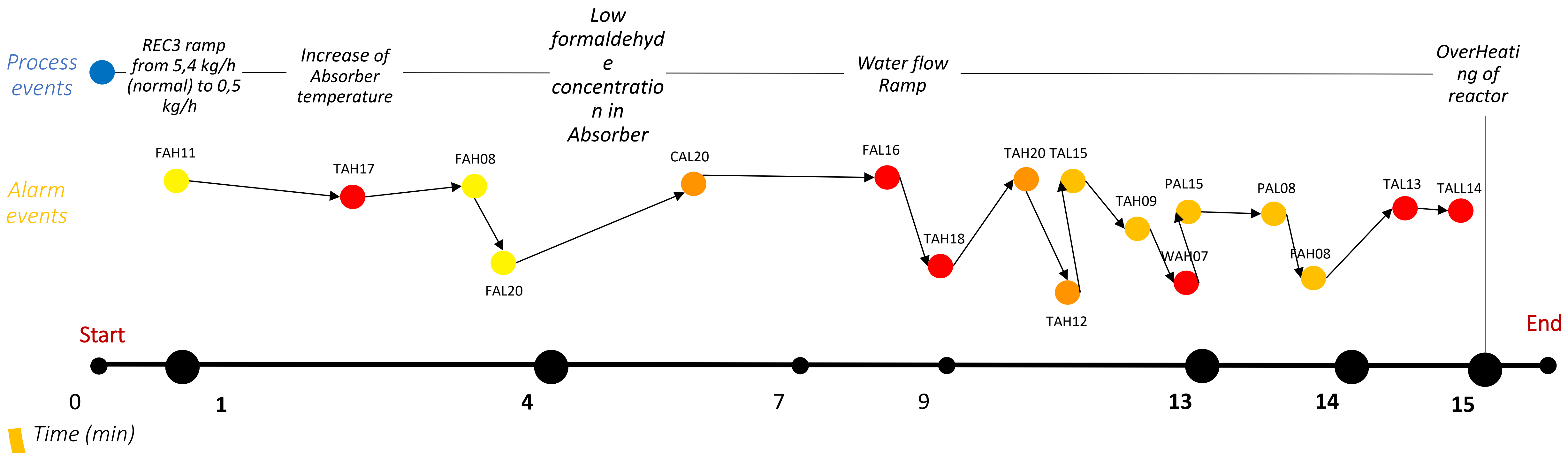




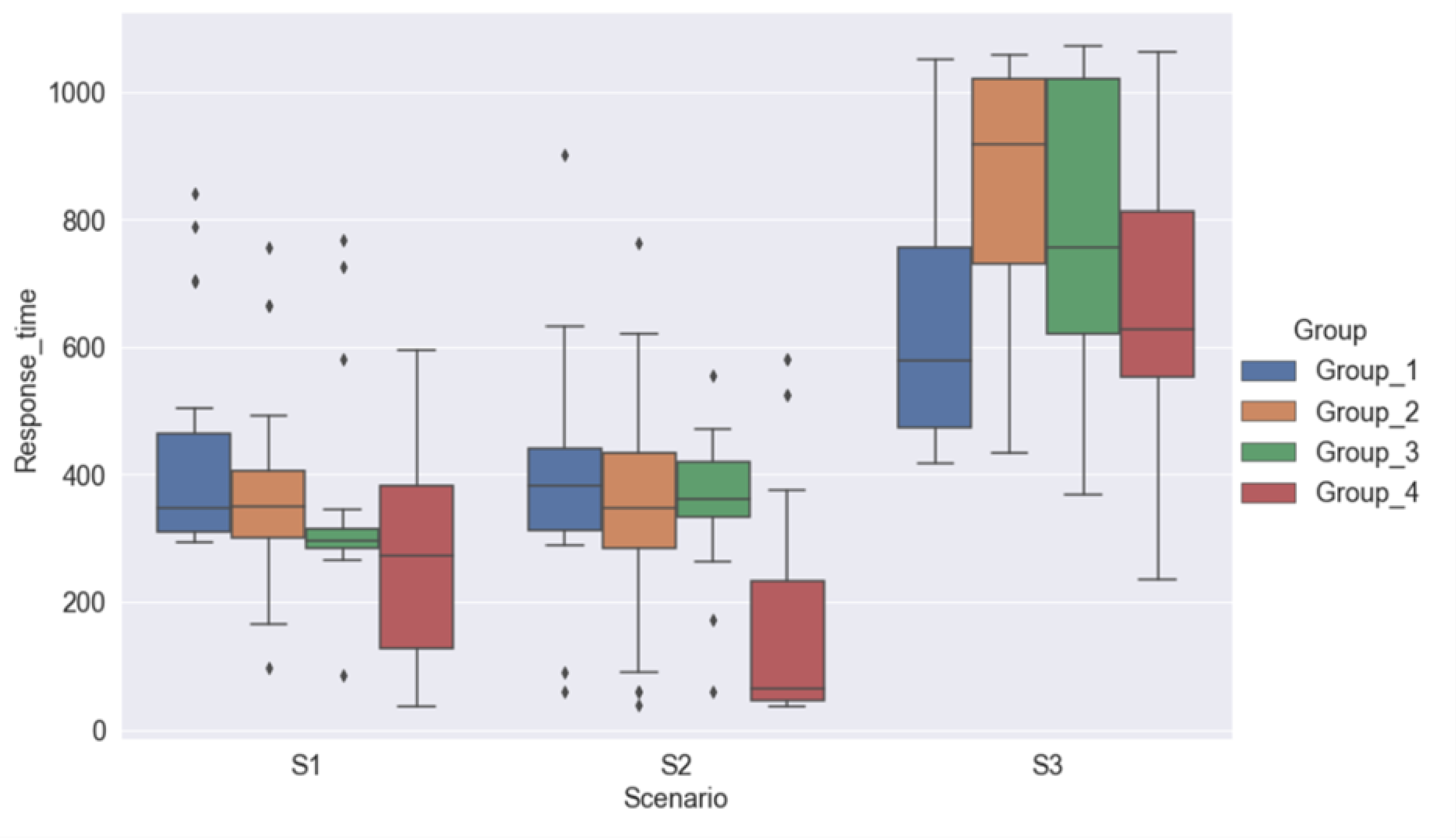
Scenario 2: Events and Task timeline



Scenario 3: Events timeline

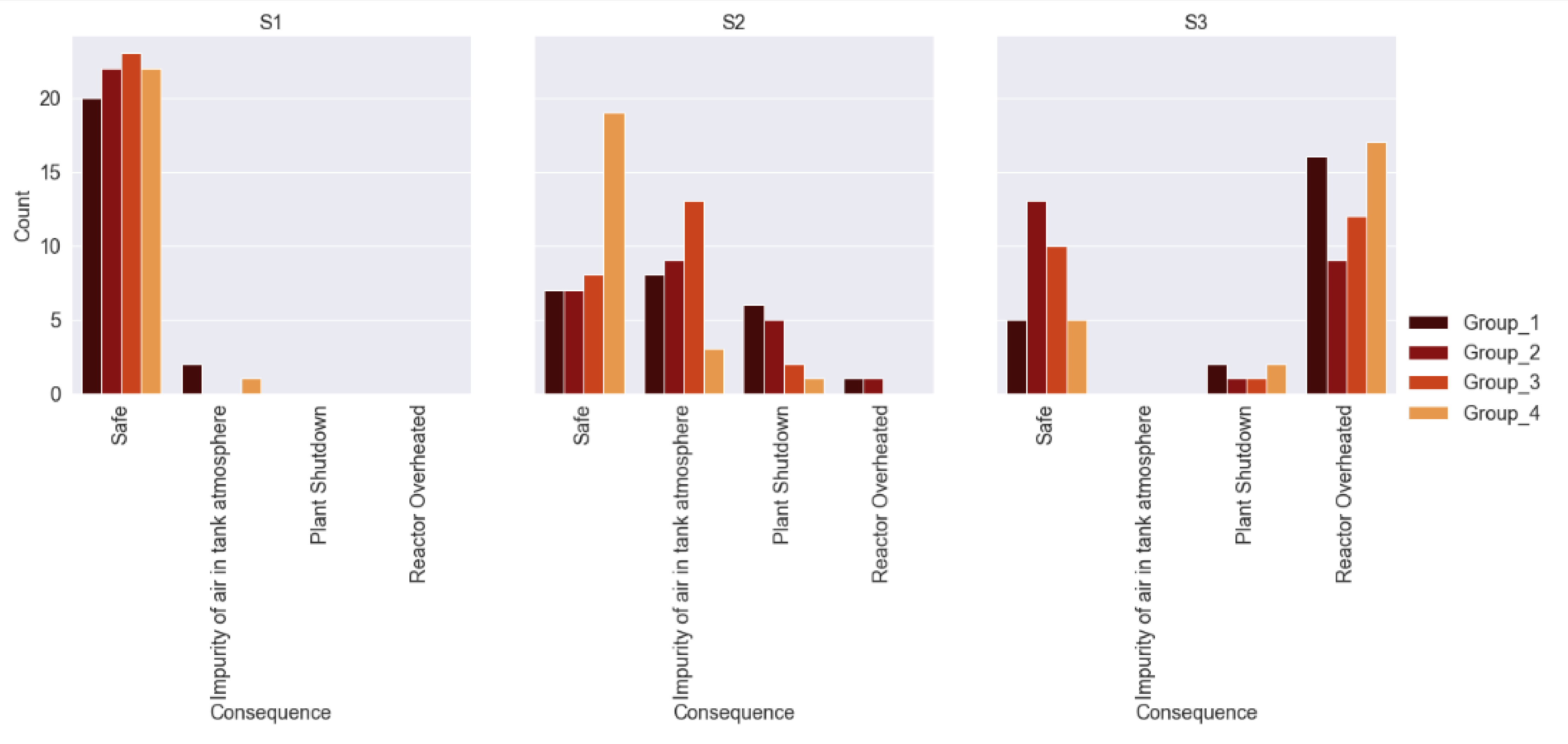


Response Time



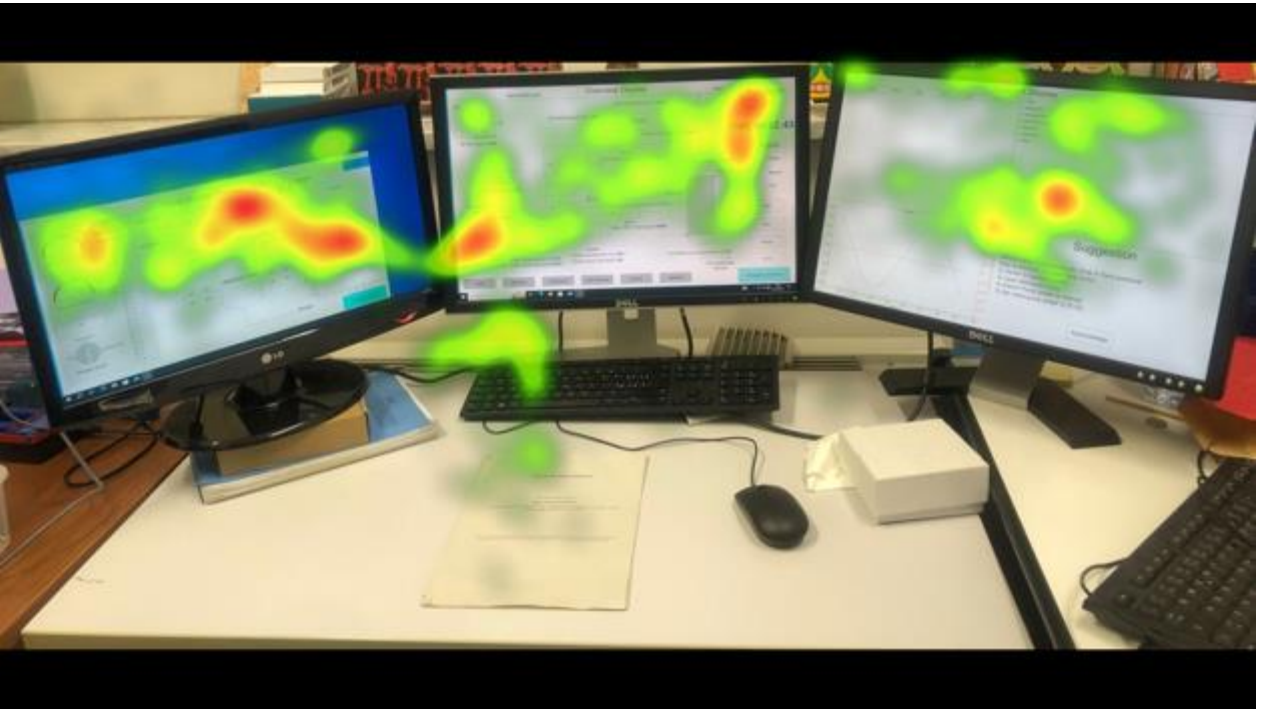
Analysis using python in jupyter notebook

Consequences



Analysis using python in jupyter notebook

Within Participants (Group 4)



Group 3 vs Group 4

Cross Analysis with Other Metrics

- It was observed that the participant that follows through AI suggestions tend to **solve** the problem **earlier** with **lesser** task **load**.

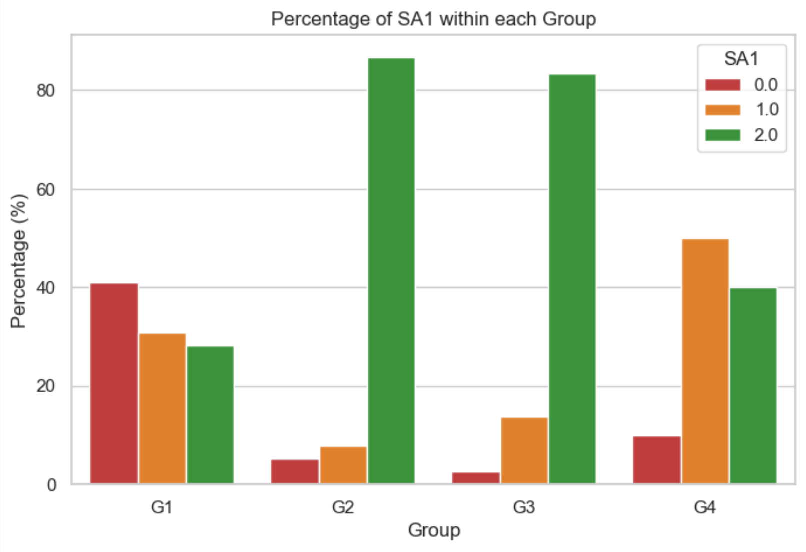
- However, with **lower** situational **awareness** as compared to the other participant that followed screen procedures.

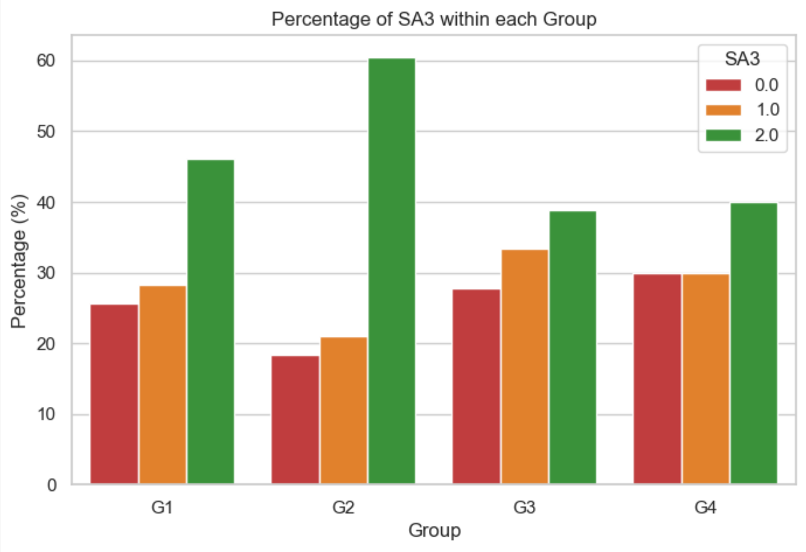
Situation Awareness Observation protocol results

SPAM-adapted Questions:

1. Which of these alarms, in your opinion, requires to be verified first and why?
2. Why do you think the critical alarm is activated? And what do you intend to do?
3. After your actions, what do you think is going to change in the system? Why?

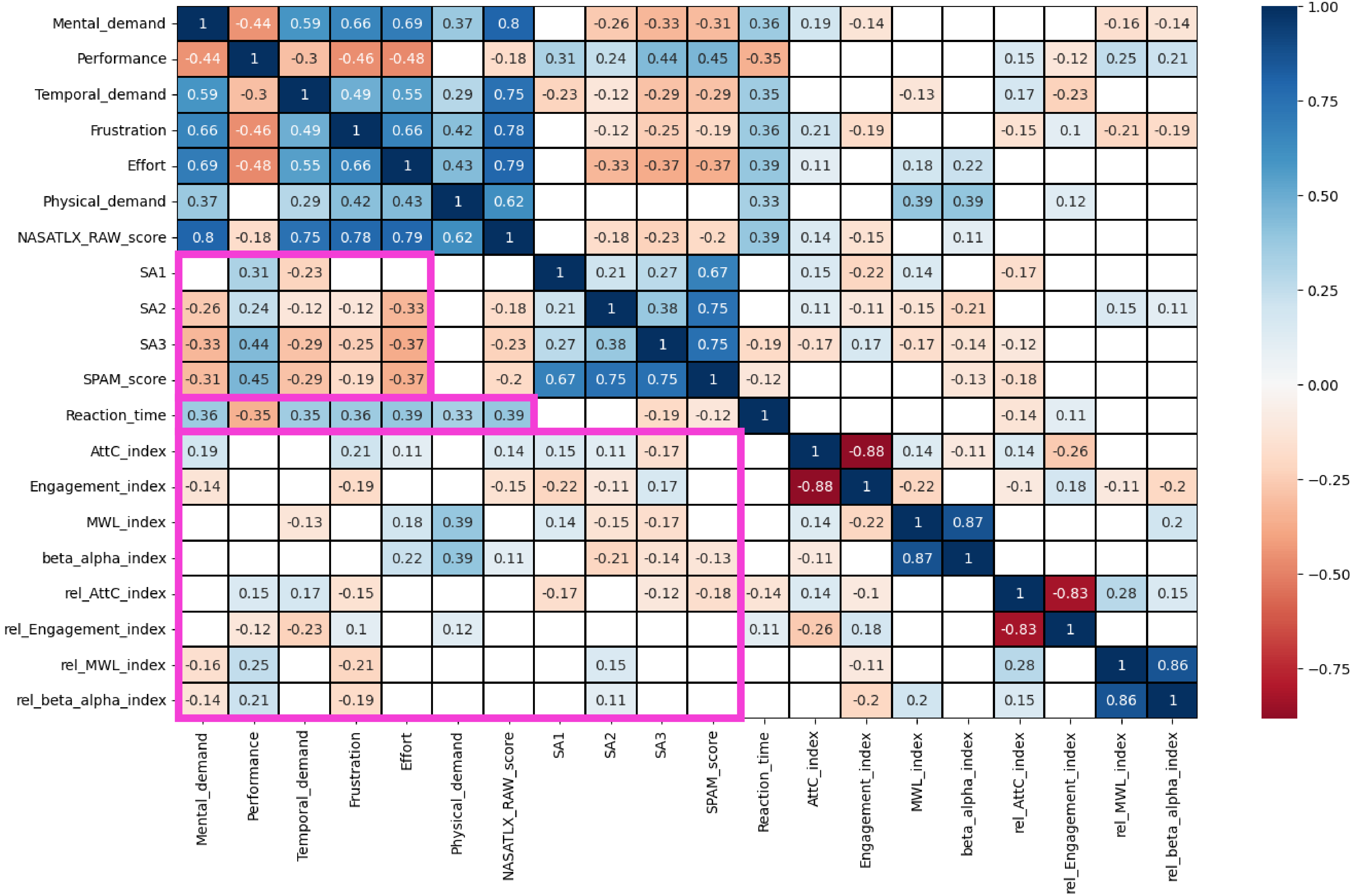
Results from S1



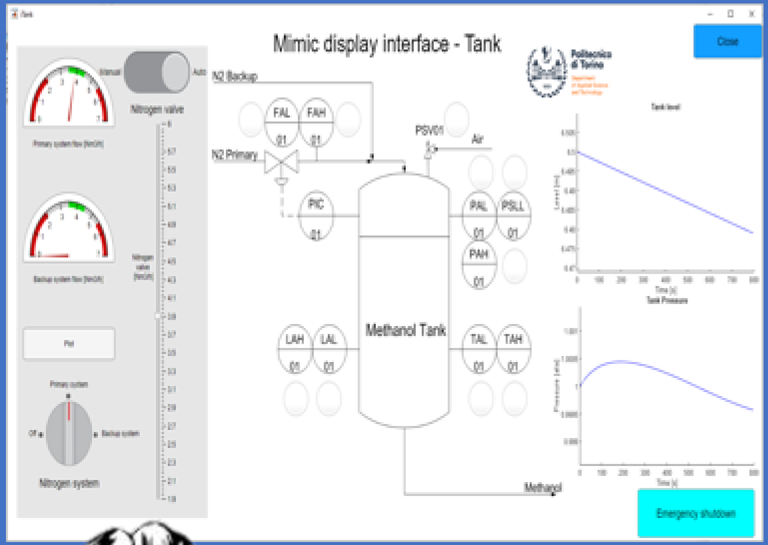


EEG correlation Matrix

Reaction time has been added to the correlation matrix. For the 5 mins Recording after main alarm for each scenario.



general observations

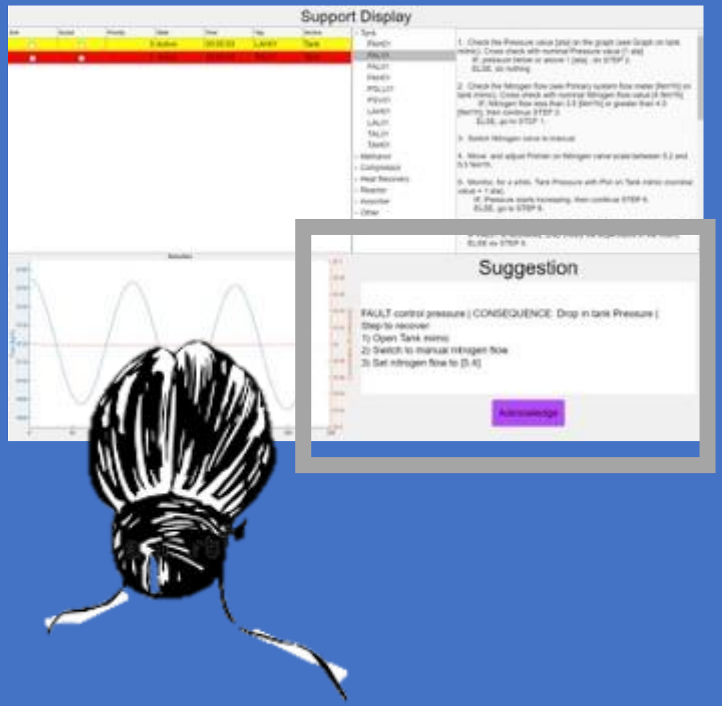


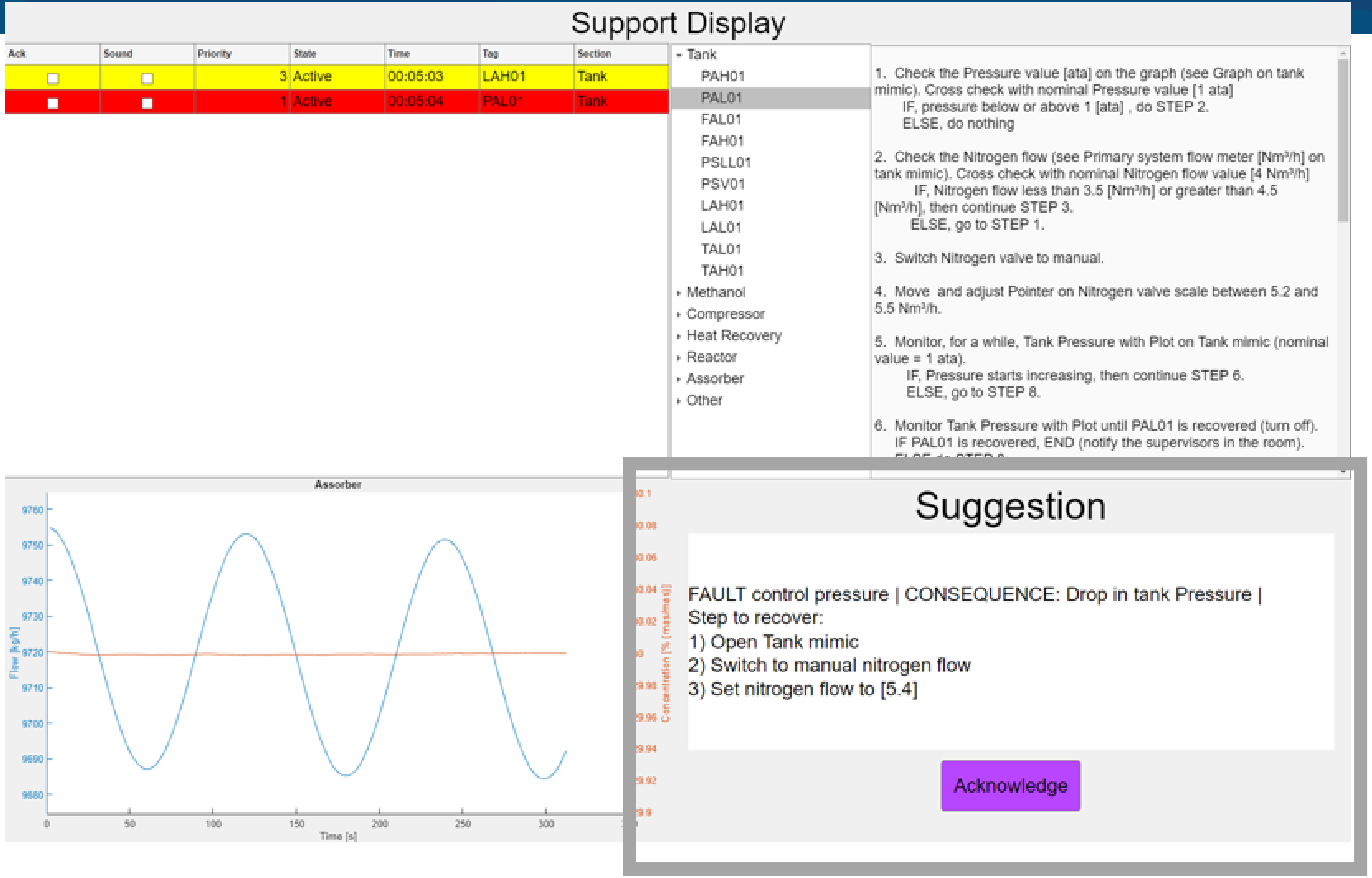
Observation	Reason	Impact	Recommendation
Unexpected behaviours: More mimics opened. More alarms were acknowledged than expected.	- Pressure - Poor alarm rationalisation – G1 (Poor awareness)	- Performance - Accuracy	Alarm prioritisation.
Unexpected behaviours: Clicking the **wrong buttons**.	Buttons had similar colours or had close proximity to each other.	Premature plant shutdown.	- Uniquely assign colours per function. - Maintain distance
Similarities in outcomes: The performance of people with paper procedures is considerably similar to that of the digital format. (Task: easy to medium complexity).	**- Simultaneous interfacing** by those in Group2.	Near performance to G2.	**- Limit scrolling** with digital interfaces. - System positioning on Head movement **- Alarm links to procedures** or other features to ease search task.

The case for Labelling historian and Log data

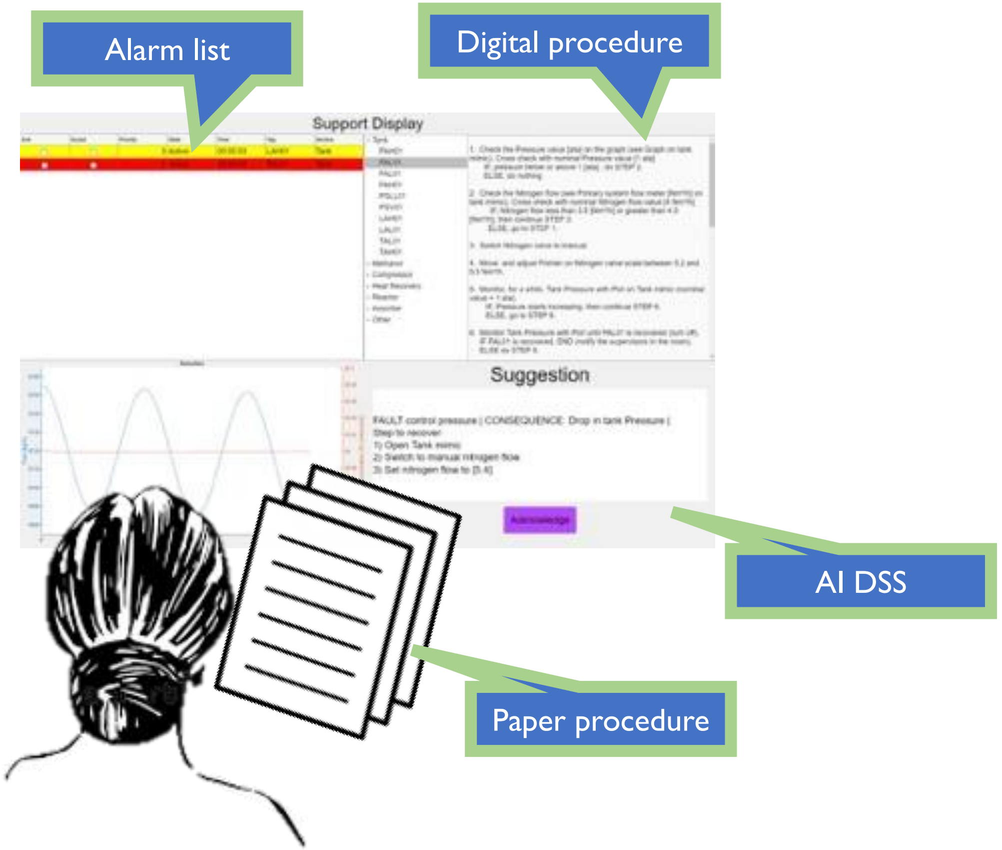
CISC
Collaborative Intelligence for Safety Critical systems

AI SUPPORT





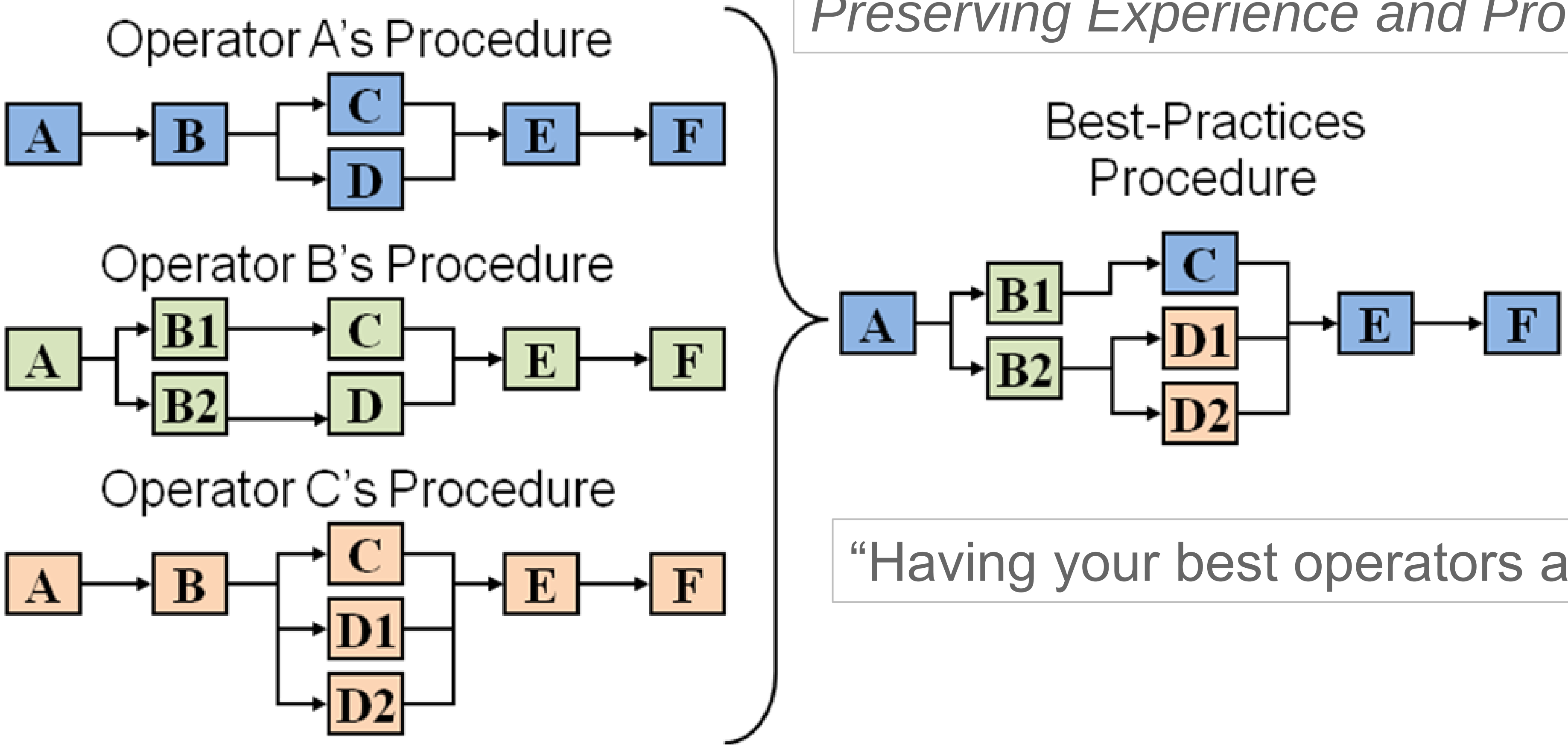
Future work



- Redesign the AI DSS Display (only for summarising situation no details instructions)
- Redesign Digital Procedure
- Real-time Operator-System Interaction Modelling
- Labelled human action in historian or DCS logs to also correlate first response action to the alarm they refer to
- Identifying other possible way to solve situations.

Human Factors in system design: ISA standards

Procedure Management (ISA 106)



Preserving Experience and Procedural Knowledge

"Having your best operators all day, every day"

Capturing Best Practices Procedures

CISC

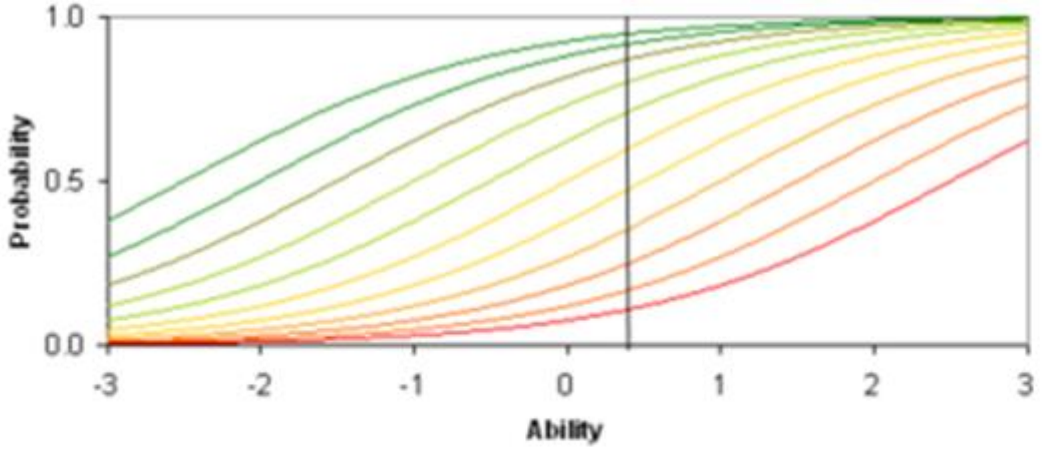
Collaborative Intelligence for Safety Critical systems

Cross sectional aspects in collaborative intelligence applications

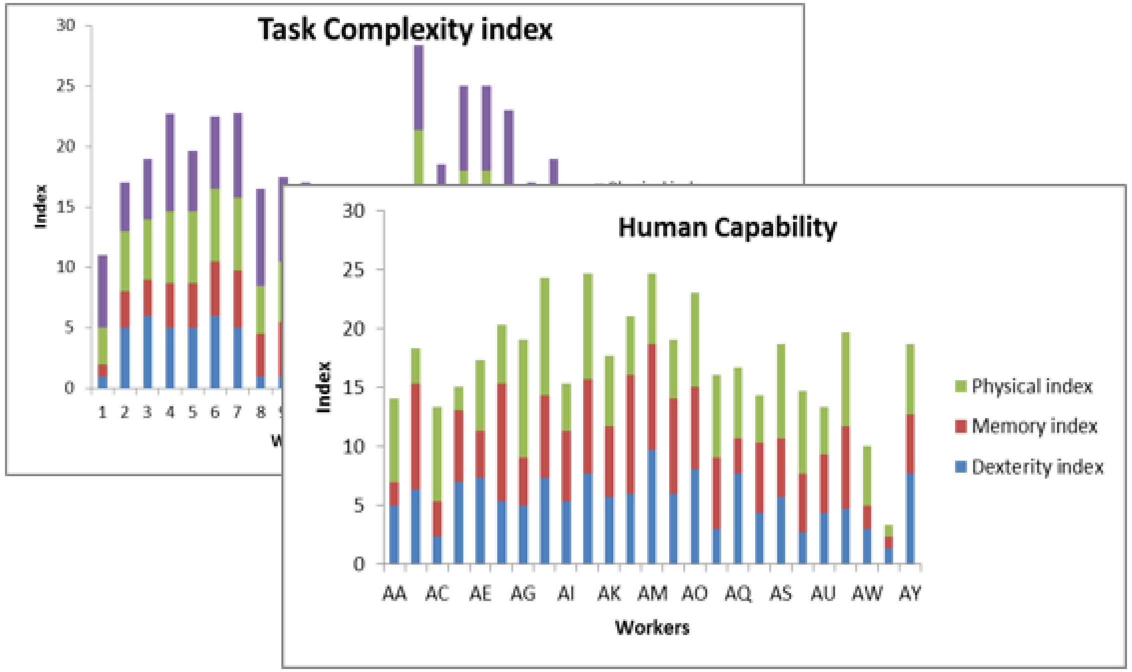
HRA models to understand humans

Mental workload is a variable closely connected with Human-System Performance. *

$$\Pr\left(X_{ni}=1\right)=\frac{e^{b_n - d_i}}{1+e^{b_n - d_i}}$$

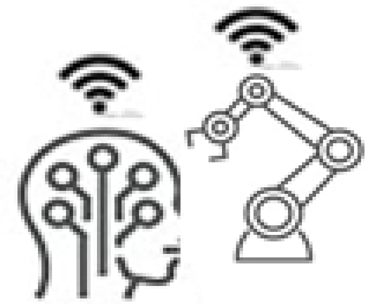


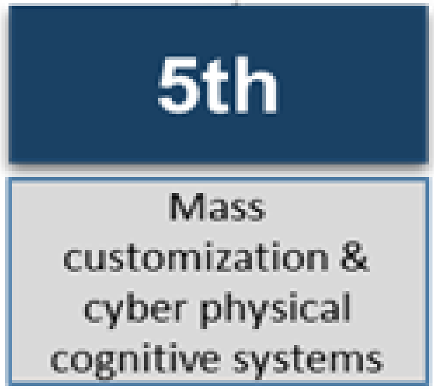
Worker performance can be individually characterised by observable characteristics some of them obtained via bio-sensors. *



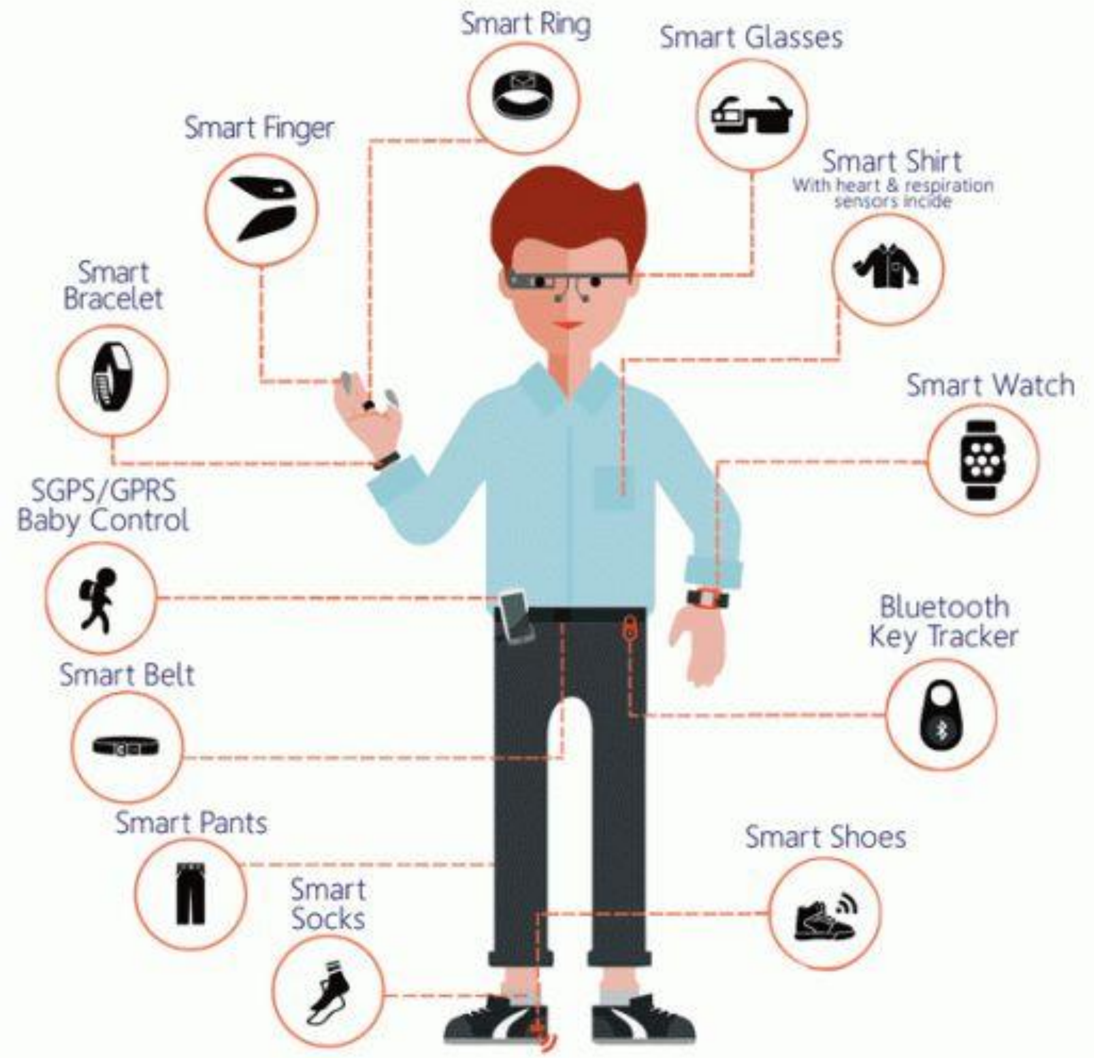
* Leva et al. “Task complexity, and operators’ capabilities as predictor of human error” in ESREL 2018

Moving towards a more pervasive assessment of the humans in the systems

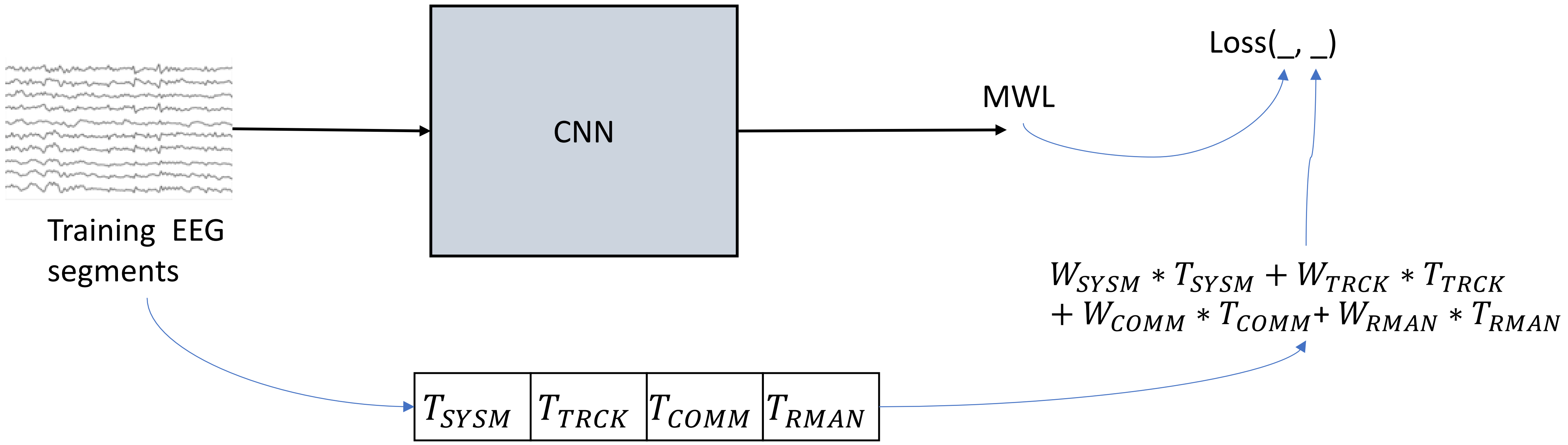




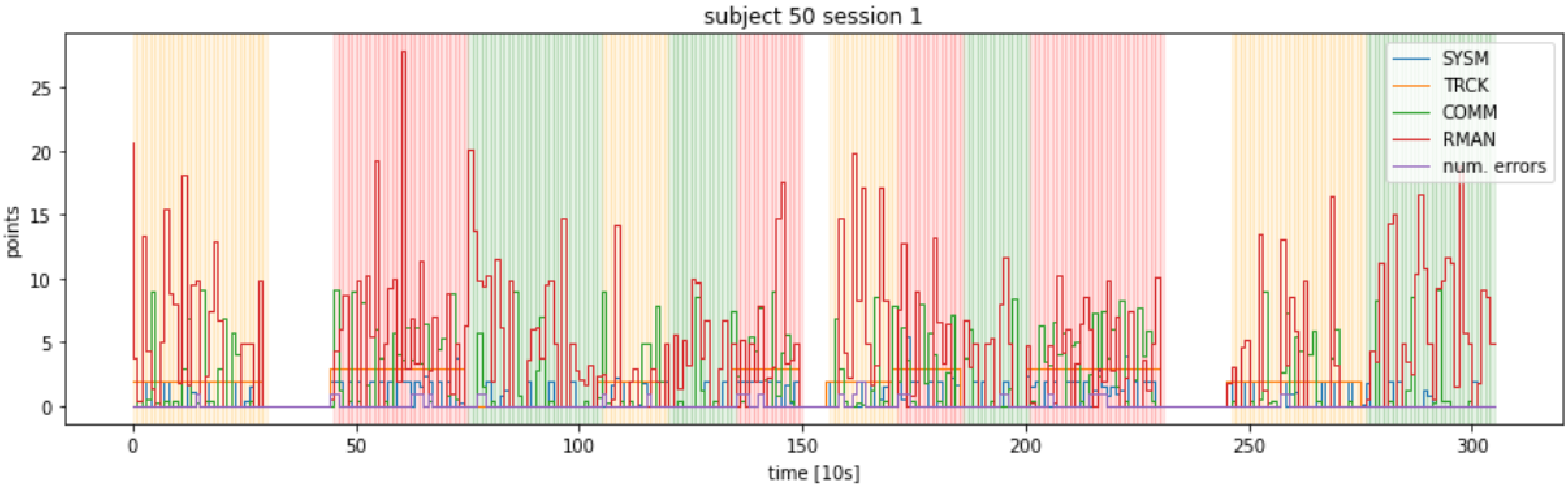
- **IOT, wearable technolgies and AI are enchancing capacity to assess the human in the system in a way that was not possible before.**



Continuous modeling of MWL (Milos)



A Deep learning approach for EEG data analysis to recognise high mental workload situations



- Instead of modeling MWL with custom tasks difficulties – make the NN learn tasks difficulties
- Problem Model that predict NASA MABT task does not perform well on assembly task

Conclusions Key challenges and opportunities

Function allocation: how can we keep situational awareness how can we deliver context awareness. . In other world REAL HITL

Inform design of the HMI and for automation so as to get the best of both worlds (fast data processing, for AI systems, leaving room for understanding and therefore use power of intuition for us" humans")

How to better support collaborative intelligence

build close feedback loop between observable variables and human performance probability estimations

Ethical: substitute versus meaningful work and task environments. Being realistic

new sources of data offered such as advances in Neuroergonomics for real time detection of changes in our conditions: mutual monitoring between human and AI

Stavanger ESREL SRA-E 2025

Home | Submissions & Deadlines | Program | Registration | Travel & Venue | Sponsors & Exhibitors | Contact

Welcome to ESREL SRA-E in Stavanger, Norway!

15-19 June 2025

ESRA European Safety and Reliability Association

HOME | NEWS & EVENTS | COMMITTEES | ESREL | JOIN ESRA | ABOUT

METHODOLOGIES

- Accident and Incident Modeling
- Decision Making under Uncertainty
- Foundations of Risk and Reliability Assessment and Management
- Human Factors and Human Reliability
- Maintenance Modeling and Applications
- Mathematical and Computational Methods in Reliability and Safety
- Organizational Factors and Safety Culture
- Prognostics and System Health Management
- Resilience Engineering
- Risk Assessment
- Risk Management
- Structural Reliability Applications
- System Reliability Applications
- Uncertainty Analysis

Human Factors and Human Reliability

The focus of this technical committee is the analysis of human performance for the safe and reliable operation of complex socio-technical systems. The technical committee fosters research and collaborations on methods, applications, and on the use of analysis results for decision-making.

This committee keeps together the human reliability analysis and human factors disciplines: the ESREL conference is one of the few occasions in which both communities meet. Our aim is to jointly benefit from sharing the latest advances of both fields.

Examples of topics of interest for the committee are:

- Characterization, measurement, and models of performance influencing factors
- Integration of the human component in risk and resilience engineering
- Human performance models and data in complex socio-technical systems
- Human reliability analysis

The committee maintains close links with the

- Human Reliability Analysis Society, http://hrasociety.org/blog/
- The Probabilistic Safety Assessment and Management (PSAM) Conference http://www.iapsam.org/

Chair:
Luca Podofillini - Paul Scherrer Institute, Switzerland

Co-Chair:
Chiara Leva - Technological University Dublin, Ireland