**Collaborative Intelligence for Safety Critical systems**
**Marie Skłodowska-Curie Actions (MSCA)**
**Innovative Training Networks (ITN)**

# Deliverable 3.2.
# Data integration need for AI in CI applications

| Deliverable information | |
|---|---|
| **Deliverable leader** | SCCH |
| **Document due date** | 31/12/2022 (M24) |
| **Actual submission date** | 25/01/2023 (M25) |
| **Version** | 1.0 |

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

| Document Control Page | |
|---|---|
| **Document Title** | D3.2. Data integration need for AI in CI applications |
| **Project title** | Collaborative Intelligence for Safety Critical systems (CISC) |
| **Coordinator institution:** | Technological University Dublin |
| **Main scientific representative:** | Dr. Maria Chiara Leva (mariachiara.leva@tudublin.ie) |
| **Project duration** | 1st January 2021 – 31st December 2024 (48 months) |
| **Website** | www.ciscproject.eu |
| **Related WP/Task** | WP 3: AI/ Machine learning for collaborative intelligence frameworks |
| **Deliverable nature** | Report |
| **Dissemination level** | Public |
| **Description** | Data integration need for Artificial Intelligence in Collaborative Intelligence applications |
| **Creation date** | 03/11/2022 |
| **Keywords:** | Data integration, Artificial Intelligence, Collaborative Intelligence |
| **Contributors** | Georgios Chasparis, Ammar Abbas, Carlos Albarrán Morillo, Devesh Jawla, Carlo Caiazzo, Joseph Mietkiewicz, Aayush Jain, Ines Fernandes, Shakra Mehak, Doaa Almhaithawi, Miloš Pušica, Naira Lopez, Houda Briwa, Chidera Amazu, Andres Alonso Perez, Hector Diego Estrada Lugo, Maria Chiara Leva |
| **Review status** | Draft / Released / Validated / **Submitted** |

## Versions

| Revision | Date | Author | Organisation | Description |
|---|---|---|---|---|
| 0.1 | 28.11.2022 | Georgios Chasparis | SCCH | Document structuring |
| 0.2 | 05.12.2022 | Carlos Albarrán Morillo | Polito | ESR 11 Contribution |
| 0.3 | 05.12.2022 | Devesh Jawla | TU Dublin | ESR2 Contribution |
| 0.4 | 07.12.2022 | Carlo Caiazzo | FINK | ESR 7 Contribution |
| 0.5 | 15.12.2022 | Joseph Mietkiewicz | HUGIN EXPERT – TU Dublin | ESR 3 contribution |
| 0.6 | 15.12.2022 | Aayush Jain | IMR | ESR 12 contribution |
| 0.7 | 15.12.2022 | Ines Fernandes | Unimi | ESR 8 contribution |

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

| | | | | |
|---|---|---|---|---|
| 0.8 | 15.12.2022 | Shakra Mehak | Pilz | ESR 13 Contribution |
| 0.9 | 16.12.2022 | Chidera Winifred Amazu | PoliTo | ESR4 Contribution |
| 0.10 | 19.12.2022 | Houda Briwa | TU Dublin | ESR1 Contribution |
| 0.11 | 19.12.2022 | Doaa Almhaithawi | MATHEMA | ESR9 Contribution |
| 0.12 | 22.12.2022. | Miloš Pušica | mBrainTrain | ESR6 Contribution |
| 0.13 | 22.12.2022 | Naira Lopez | DIGITALSME | ESR10 Contribution |
| 0.14 | 22.12.2022 | Ammar Abbas | SCCH | ESR5 Contribution |
| 0.15 | 23.12.2022 | Andres Alonso Perez | mBrainTrain | ESR14 Contribution |
| 0.16 | 26.12.2022 | Ammar Abbas | SCCH | Finalisation and formatting |
| 1.0 | 23.01.2023 | Georgios Chasparis | SCCH | Reviewing and final version |
| 1.1 | 25.01.2023 | Naira Lopez, Diego Estrada | DIGITALSME, TU Dublin | ESR10 contribution, final revision for submission |

# Table of contents



This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

# Executive summary

The present document constitutes Deliverable D3.2, with title “Data integration need for AI in Collaborative Intelligence applications” in the framework of WP3 “AI/ Machine learning for collaborative intelligence frameworks”, and reports on tasks “T3.2: Definition of end-user potential need from within and outside the consortium”; “T3.3: Categorization of current issues, accidents and shortcomings stemming from the overlooking of AI & human integration for normal or abnormal operations”; “T3.4: Available procedure and applications from academic, industrial and infrastructure domain. Shortcoming and development needs”. This deliverable further contains the updated description of each LiveLab followed by an update in the structure of the data pipelines, the proposed methods for data collection and the purposes for which they will be used as well as the ethics related to each case. This deliverable has been edited by the 14 Early-Stage Researchers and led by SCCH beneficiary and the coordinating board.

# 1. Introduction

In this deliverable and in the following paragraphs, we will provide a short introduction of the considered LiveLabs, while also discussing the involved data processes and necessary data integration. Our goal is to address the challenges involved in the context of data requirements for better designing human-machine collaboration in industrial processes, or, briefly, *collaborative intelligence*.

**LiveLab 1.**

The current technology available for robotic solutions that can work collaboratively with human workers by sharing their skills is entering the market. This has become a new frontier in industrial robotics that allows combining the advantages of robots, which enjoy high levels of accuracy, speed and repeatability, with the flexibility and cognitive skills of human workers. However, to achieve an efficient human–robot collaboration, several challenges need to be tackled. First, this LiveLab will focus on the analysis of the physiological state of the operator during the collaborative tasks through advanced neuroergonomics experiments. The proposed modular workstation allows to implement data coming from electroencephalograms (EEG), electromyography (EMG) sensors and collaborative devices such as cobots and\or poka-yoke systems to analyse the physiological state of the operator during collaborative tasks.

Additionally, this LiveLab will be generating and implementing data for the study the factors that affect the performance of operator teleoperating of robots. The challenge lays in the reduced operator’s situational awareness and lack of information to make sound decisions. Therefore, within this LiveLab datasets to assess states of degraded performance will be collected by using wearable sensor technology. The first task consists of the identification and validation of different telerobot interface factors and how they affect the operator internal state and task performance, and then the creation of a dataset by collecting multi-modal data from participants while they perform a real telerobot task and while different performance-related internal states are evoked and assessed. These systems will collect physiological signals to train a deep learning model to predict performance-related operator’s states.

To ensure a higher interconnectivity between the human and robot, it is important that interactive programming methods for robots are properly designed, so that human operators can easily program and interact with the robot. To develop an interactive programming method for robots, an Interactive Task Learning (ITL) framework has been proposed which transfers human skills to robots through

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

demonstration. This system is supported by a deep neural-network-based anomaly/failure detection module and it is deployed as a safeguard to prompt a user in safety-critical events.

Finally, safe interaction must be guaranteed to prevent harming of the human operators and ensure the reliability of the robotic systems. For this reason, this LiveLab will also be focused on the safety of the operator in task demonstration as well as the assessment of human performance in kinaesthetic teaching. This part of the project will gather the necessary data to work on the risk assessment paradigm which addressed the use of hybrid standardization format and focuses on the functional safety of cobot cell.

**LiveLab 2.**

The main focus of this LiveLab is on the manufacturing operations in a large-automotive plant. The reason for focusing on this specific industry relies on the notion that the manufacturing sector is still widely based on human operations. They still play a crucial role in many aspects of the industry. This is because manufacturing processes often require a high degree of flexibility and adaptability, which can be difficult for machines to replicate. Thus, the main need for data for this LiveLab will come from the operator's human performance in the manufacturing industry. For this reason, to take full advantage of the capabilities of human operators and robots, the ESRs involved in LiveLab 2 will need sensory data from the machinery interaction (such as welding operations) performance will also be collected (parts completed without errors, unsafe conditions etc.) and environmental conditions. The data regarding the performance will be collected from the operators themselves, such as physiological parameters or human errors performed during the task execution, to assess the capabilities of the operators to perform the given task in a real-time.

In addition to the operator performance datasets, the collected data at IVECO will be IoT and sensor data, these will be time series data which will need to be pre-processed. For IoT and Sensor Time-Series Data, anomaly detection uses AL, through uncertainty sampling. For example, suppose a highly imbalanced data with 999 regular events and only one abnormal event. In that case, the algorithm provides a label for this one abnormal event. There are usually only very few key events that require human attention in an exceptionally long time series. And therefore, to train a data driven anomaly detection model it may be necessary to review and label an exceptionally enormous amount of data points to identify enough anomalous datapoints for the data driven anomaly detection model to be trained on. This requires a human in the loop to teach and verify the ML algorithm for robustness, this kind of paradigm is known as Active learning.

**LiveLab 3.**

Operators in control rooms of complex socio-technical systems like process plants are faced with increased cognitive loads from their current process control configurations. This is because they are left with tasks that involve a lot more information processing and demand great attention with less manual or physical roles. However, despite the emphasis on human error as a leading contributor to safety, the risks and uncertainties from such evolving processes or monitoring capabilities of the operators and the related influencing factors are usually not considered during safety analysis. Therefore, an enhanced approach and tools are needed to capture, monitor, and address risks and uncertainties. This work aims to discuss a risk-based decision-making framework and tools that

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

integrate human, organisational and technical dynamics, together with some tools for process control optimisation and management. Monitoring the operators' dynamics on a task given pre-identified process safety scenarios is a way to reduce the uncertainties in the prediction of control room operators' actions. For this work, process safety events have been identified from a simulated interface of a formaldehyde production plant. Experiments, with and without support systems, would be set up using students and potentially control room operators to simulate process operations and collect relevant data. The operators' psycho-physiological data would be collected with qualitative data from surveys, audio recordings, and process data from the simulated plant. Furthermore, data analysis and machine learning techniques will be used to analyse and integrate the collected data for human error probability and uncertainty estimation. This data can serve as input to estimate the human error probabilities as influenced by human, organisational and technical factors. Furthermore, the aim is to provide an online decision support tool for the operator. The model takes as input the process values to deduce the state of the system. The different states are related to different risks. According to this state, the influence diagram recommends to the operator the action, defined by an expert, that minimises the cost, given a monetization of the relevant risk and considering the error rates. In addition, reinforcement learning is used to specify the action by providing the best possible value to set.

**Legal and Ethical Implications for each LiveLab.**

For several reasons, it is essential to vet the sensor data collected from machines or humans carefully. First, there are ethical and legal considerations that need to be taken into account when collecting and using sensor data. For example, the data must be collected and used to respect individuals' privacy and comply with relevant laws and regulations. To this end, for each LiveLab a list of potential risks to look out for will be provided and updated, so as to allow for a more dynamic, adjustable and straight-forward analysis of considerations to bear in mind. Once the risks have been laid out, it makes sense to propose some of the most relevant recommendations offered by the literature that can help answer them.

This document "D3.2. Data integration need for AI in CI applciatins", contains the description of the data pipelines that will be produced in each LiveLab, the experiments developed to obtain the data and the purposes for which they will be used as well as the Ethics around each case. The next section will describe briefly each of the Live Labs. Then, in Section 3, the contribution from each ESR is contained and sorted by Live Lab. Finally, as this is a work in progress at the time of writing this deliverable, some preliminary conclusions for the three LiveLabs are provided followed by the immediate future actions.

# 2. Brief description of Live-Labs

Within the framework of the CISC project three Live Labs are proposed, a detailed description of which has been provided in deliverable D5.1 (link). In this section, we also provide a brief summary of the live labs as well as the involved use-cases targeted within this project by the ESRs.

## Live-Lab 1: Human-Robot Collaboration

Live Lab 1 is divided into three different scenarios located in IMR's pilot factory in Mullingar and in the Faculty of Engineering, University of Kragujevac. This Live Lab will focus on how operators can better interact with robots at various control levels. The efficiency and safety of these operations will be studied and data concerning human factors, such as attention and comfort etc., collected.

*Live-Lab 1.1. (ESR#7 Carlo Caiazzo)* Human ergonomics and task partitioning in human robot collaboration tasks

The case study investigated in the Live Lab 1.1. at the Faculty of Engineering, University of Kragujevac (FINK), Serbia, proposes a hybrid industrial modular workstation where it is possible to conduct advanced neuroergonomics experiments. The proposed modular workstation allows to implement various systems such as EEG, EMG sensors and collaborative devices such as cobots and\or poka-yoke systems to analyse the physiological state of the operator during collaborative tasks.

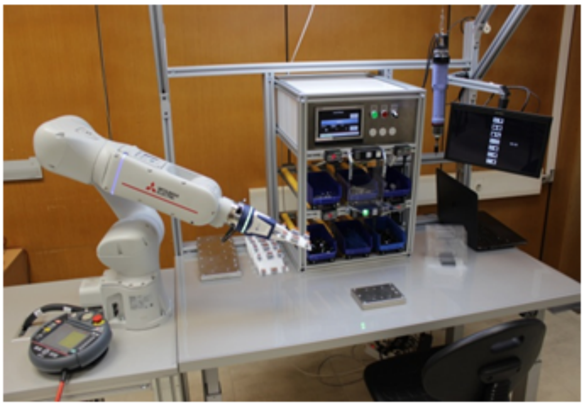


Fig.2.1.1: proposed industrial hybrid workstation in Live Lab 1 at FINK

*Live-Lab 1.2. (ESR#8 Ines Ramos)* Human performance monitoring for human-in-the-loop telerobot operations

With the recent advances in wearable technology, it is now possible to monitor the operator's internal state, through changes in their physiological signals. Intelligence systems can then be made aware of the operator's state, not only to avoid critical situations of degraded performance, but to act proactively and adapt the automation level, the interface or interaction mode to the operator's needs to achieve optimal system performance.

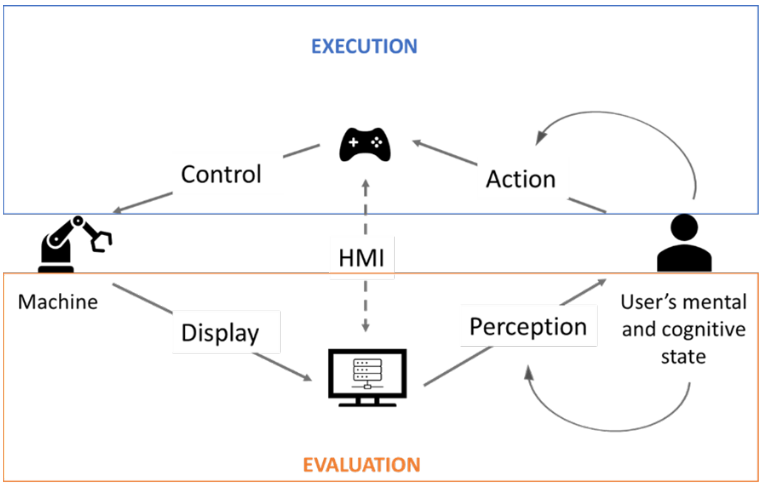
In teleoperation of robots, where the operator is remotely located, the challenge lays in the reduced operator's situational awareness and lack of information to make sound decisions. This is therefore an appropriate case study to assess states of degraded performance using wearable sensor technology and collect physiological signals to train a deep learning model to predict performance-related operator's states.

This Live Lab is conducted at IMR's facilities in Mullingar in collaboration with another EU's Horizon 2020 project FineTETHER, Fine ManipulaTion in MEdical Device ManufacTuring witH TelERobotics, that studies the use of robotic teleoperation to perform complex tasks requiring fine manipulation, high performance, sterile environments, and handling of miniature components, in medical device manufacturing applications. The project has two industry partners from the medical device manufacturing industry. In both use cases, the partners are interested in teleoperation to increase yield rate of product and operator ergonomics. The medical device manufacturing industry requires

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

precise assembly tasks and the manipulation of complex objects. These operations are difficult to accomplish and program using standard robotic tools. Consequently, many medical device companies struggle to locate system integrators willing to automate a task for a reasonable budget. Additionally, since many materials are deformable, task parameters vary, leading to automated solutions which are inflexible and rely on complex jigs and fixtures. The resulting solutions are applicable to a narrow class of assembly tasks. Maintaining this type of manufacturing in a high-cost society has become a priority and thus there is an urgent need for a class of adaptable automated solutions. A human-in-the-loop tele-operation system provides a solution for these automation needs, combining the benefits of the robotic system (motion scaling for highly precise movements, speed, accuracy, repeatability, adaptable vision system) with the expertise of the operators, reducing automation costs, lowering safety requirements compared to the use of cobots or manual operations, and reducing the costs of maintaining a clean room, a common requirement for the manufacturing of medical devices. At this level of robot autonomy, the human is responsible for sensing, planning and acting, however the robot can assist with action implementation. In this particular manufacturing use-case there is no major safety concerns, however parallels can be drawn with other safety-critical applications, such as the use of telerobots for surgery, for military missions or rescue activities.

The CISC and FineTHETHER project have three shared goals for the Live Lab: improve the human-machine interface (HMI), monitor, and improve operator's performance and reduce operator's training time. The Live Lab will follow the experimental hypothesis that the human-machine interface and haptic and visual interaction factors can affect the cognitive/mental state of the operator and consequently the teleoperation task performance (schema illustrating the experimental hypothesis shown in Figure 1.2.1). A pilot study will be conducted first to identify and validate the HMI conditions and assistive features that affect the operator's internal state and performance. The next experiments' objective is then to use the designed interface conditions, validated in the pilot study, to induce performance-related internal states, such as mind wandering, effort withdrawal, perseveration, inattentional blindness and deafness and modulate task performance, while the operator is performing a fine manipulation task and their physiological signals are being monitored by sensor devices (portable EEG device from mBrainTrain and eye-tracking using Tobii pro device installed in the interface screen). The induced states will be validated with subjective measures gathered through questionnaires and objective metrics/indicators computed from the recorded physiological signals and task-related measures.




This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Figure 1.2.1 – Schema of Human-machine interaction action cycle that is at foundation of the experimental hypothesis. In human-machine interaction there is an information processing and action loop where the human operator is constantly performing actions to control the machine through the interface and receiving feedback through it also. The connection between the human intentions and the machine's behavior is the HMI, that when ill-designed can cause problems in execution, derived from "the difference between what the user wants to do and what can actually be done using controls that are available", or in evaluation, due to "the mismatch between the user's intention and expectation and the actual state of the system". Adapted from [https://core.ac.uk/download/pdf/198044413.pdf.](https://core.ac.uk/download/pdf/198044413.pdf)

The case study goal is two-fold: the first is to identify and validate different telerobot interface factors and how they affect the operator internal state and task performance, and the second is to create a dataset by collecting multi-modal data from participants while they perform a real telerobot task and while different performance-related internal states are evoked and assessed. The dataset will then be used to train a deep learning model to differentiate between these hidden internal states using multi-modal body-signals, with the goal to build a model that can be used, for example, to provide control input for real-time interface adaptation (adapt haptic or visual interface to the operator state), real-time system autonomy level adaptation, task allocation adaptation or during the interface design process of a system.

The current telebot system architecture consists of a multi-modal sensor network that ensures the information flow between the human operator (master) and the remote robotic cell (slave). From the operator's side, data about the input controllers, such as end-effector velocity control (linear and angular), robot forces trough the controller, and additional end-effector specific commands (e.g. open/close gripper) are sent to the robot. As shown in the right side of Figure 1.2.2, the cell has two robotic arms integrated for direct dual-arm bi-lateral teleoperation control, a KUKA KR-Agilus robot and a Mecademic Meca 500 robot.

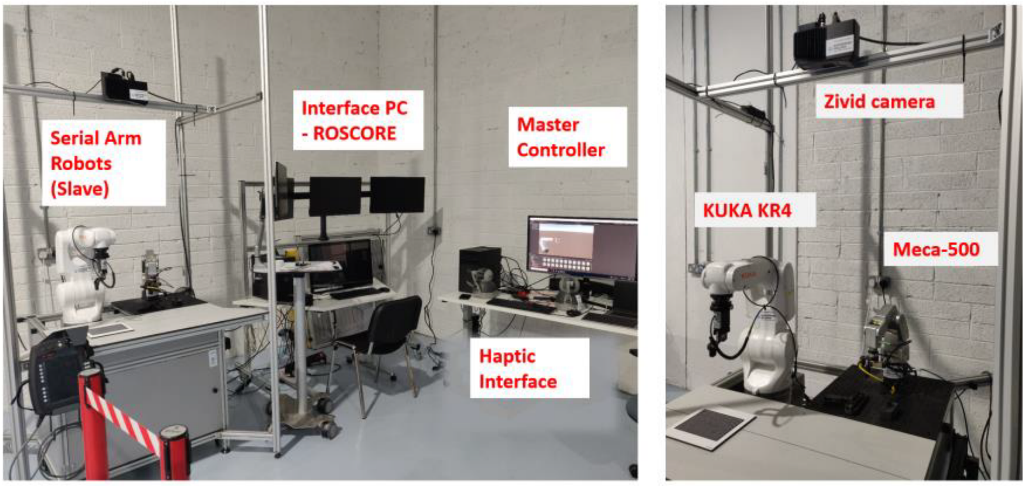



Figure 1.2.2 – Experimental set-up. Two robot arms, the KUKA KR4 and the Meca-500 (image in the right) will be used for fine manipulation. Two controllers with haptic feedback will be used to control the robot arms and visual interface will display a model of both arms, visual feedback from several cameras and other real-time information relevant to the operator.

The operator can perceive the remote environment through the visual and haptic interfaces of the system (Figure 1.2.3), receiving 2D video stream of the environment from multiple camera viewpoints, interacting with a 3D digital version of the environment (using the software Unity3D) and receiving force and torque feedback form the robot end-effector (haptic feedback).

This project has received funding from the European
Union's Horizon 2020- MSCA-ITN-2020 under grant
agreement No. 955901

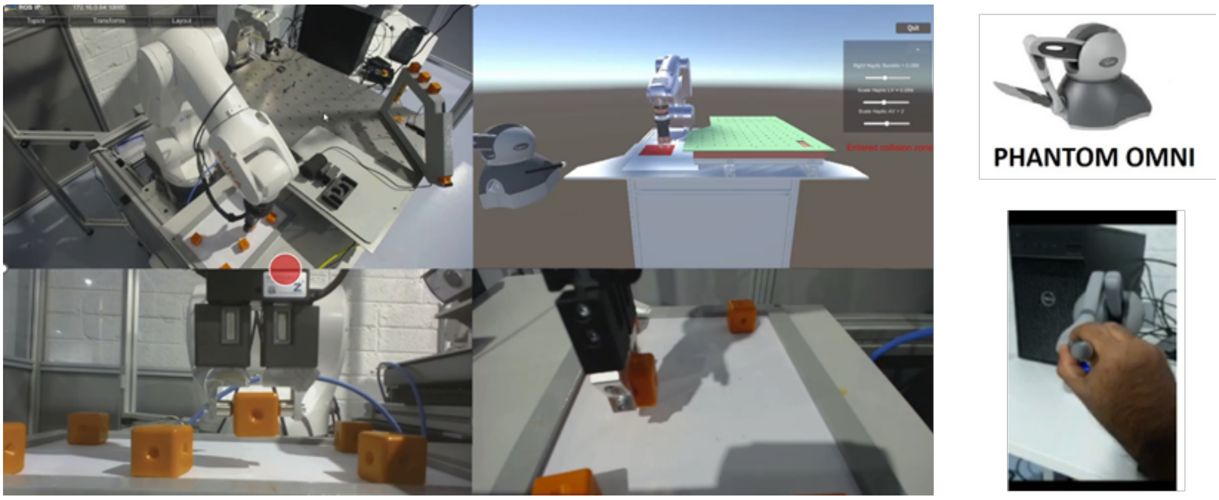



Figure 1.2.3 – Visual and haptic user interface of the telebot system. The basic visual interface (image on the left) consists of a 4-window screen with 3 windows dedicated to the real-time camera streams from 3 different viewpoints and a window for the 3D digital environment depicting in real-time the movement of the robot and offering interactive features. The haptic interface (image on the left) consists of two Phantom Omni haptic controllers that allow for translation and rotation input movements.

*Live-Lab 1.3. (ESR#12 Aayush Jain)* Programming from demonstration for fast robot programming. Human assistance during failure detection.

Early technological advances in Industry 4.0 focused primarily on improving the efficiency, quality, and productivity of an industrial cell through automation and interconnectivity. However, the development of a human-centric automation system, which supports and augments line workers, has been overlooked. In the current paradigm of system integration, an automation engineer translates the specification of a domain expert into a robot program (S. Berger et. al., 2022). However, line workers with similar domain expertise do not have the skill to program the robots. Moreover, programming methods like lead-through, walk-through, and offline robot programming interfaces are tedious and require programming expertise (V. Villani et. al. 2018). In addition, an automation engineer has to analytically decompose each task and pre-program, making the system less adaptable. Therefore, more intuitive methods of programming are desirable for instance by demonstration. Programming from demonstration (PbD) has been established as a promising method for Interactive Task Learning (ITL) (J. E. Laird et. al. 2017).

To develop an interactive programming method for robots, an ITL framework has been proposed which transfers human skills to robots through demonstration. The proposed framework, as shown in fig. 1.3.a, use a single demonstration to generate a reactive task policy using behavior trees (BTs), which are embedded with movement primitives capable of real-time adaptation. Furthermore, a deep neural-network-based anomaly/failure detection module is deployed as a safeguard to prompt a user in safety-critical events. Subsequently, we incrementally refine the system's skill library by online learning from successful and failed executions.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

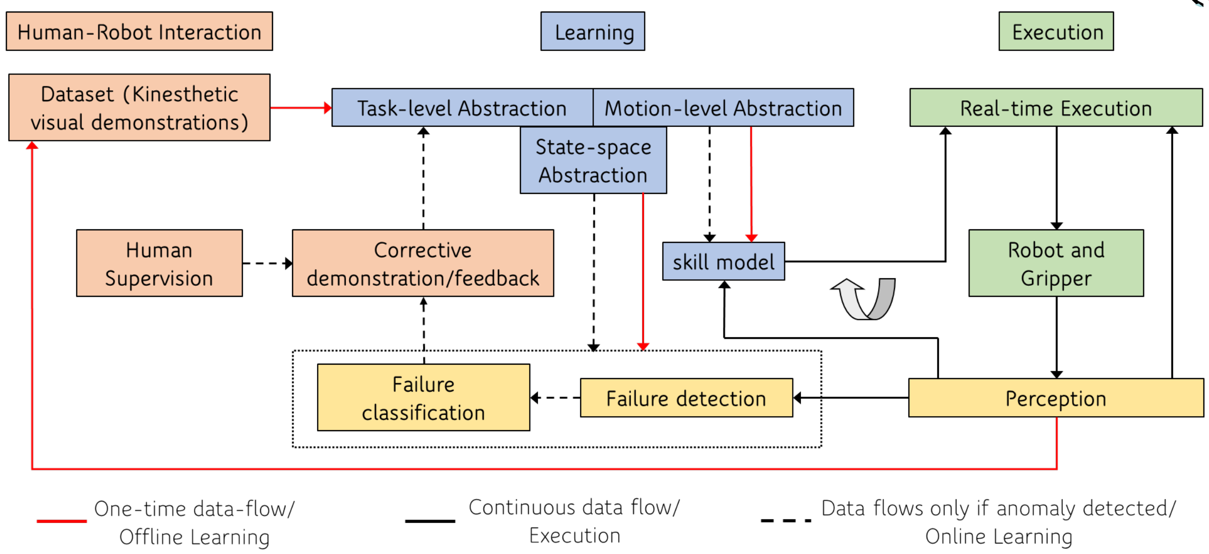



Figure 1.3.a: An overview of the proposed Interactive Task Learning Framework for Human-Robot Interaction. Different components of each module are represented in the same colour. Arrows show the stream of data flow at each phase– Offline learning (red lines), Execution phase (solid black line) and Online learning from failures (dotted black line). Curved arrows represent the learning from successful demonstration loop.

In IMR, the UR10 robot arm equipped with a gravity compensation controller, as shown in fig.1.3.b, is used for this study. Gravity compensation enables us to use this collaborative robot for kinesthetic tasks teaching the task from human demonstrations. The robot is equipped with a Robotiq 3-Finger Adaptive Robot Gripper with continuous grasp capabilities. A Microsoft Kinect v2 RGB-D camera is used as a vision sensor that provides information about the workspace and tracks the object's current state on the table. Factory set-up for testing and validation that emulates one-off manual tasks will be designed using this robot.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

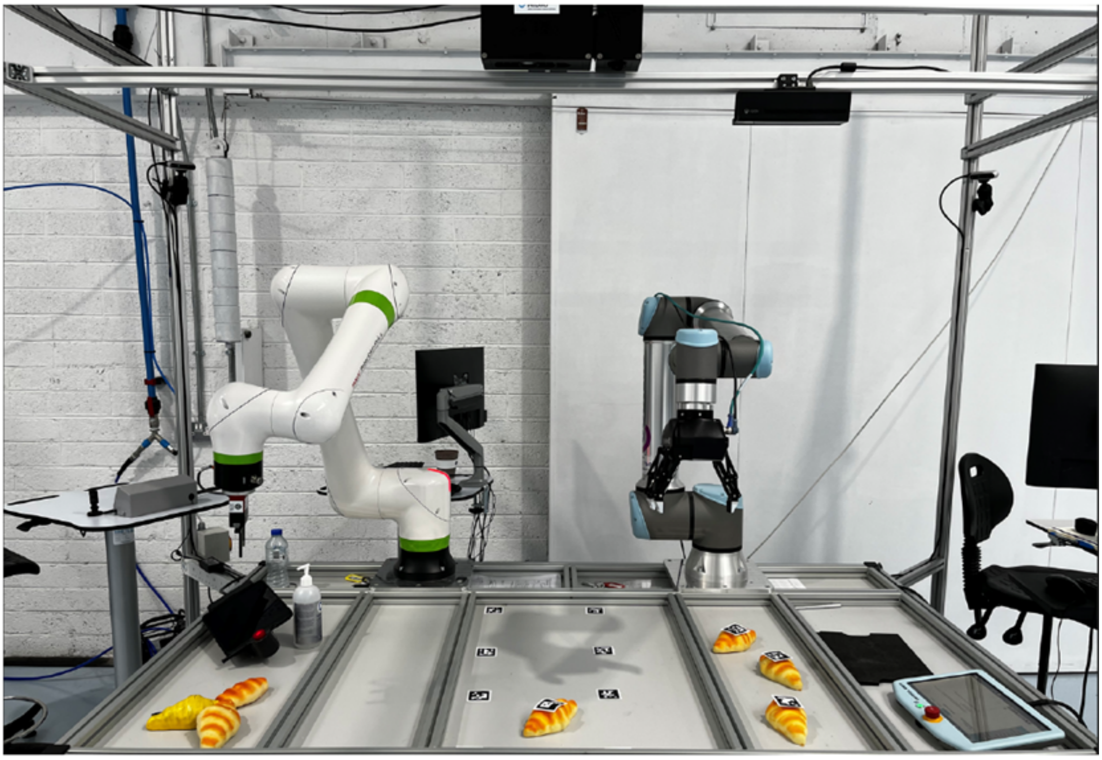


Figure 1.3.b: Livelab 1.3: Direct human robot interaction cell, featuring two collaborative manipulators- FANUC CRx10 and UR10e each equipped with grippers, and a top down RGB-D camera.

*Live-Lab 1.4. (ESR#13 Shakra Mehak)* Safety and certification during human demonstrations and corrections of robot tasks and human performance assessment in Kinaesthetic task demonstration

The use of collaborative robots in manufacturing has the potential to improve ergonomics on factory floors and increase flexibility in production. However, current human-robot collaboration (HRC) technologies are limited in terms of safety and reliability, which are crucial in risk-critical environments. Confusion about European safety regulations has led to situations where collaborative robots are used behind barriers, which negates their benefits and reduces overall productivity(Cuninka & Strémy, 2015). Developing safe and effective HRC systems for complex industrial or daily tasks remains a challenge. Factors that should be considered in order to make HRC safe and reliable include the design and layout of the workspace(Michalos et al., 2015), the capabilities of the robot, the training and supervision of human workers(Matheson et al., 2019), the communication and coordination between humans and robots(Hiatt et al., 2017), and the development and implementation of safety standards and regulations. In order to enable the adoption of HRC technology with confidence, there is a need for harmonized European norms related to HRC. Assessing the safety of human-robot collaboration is an ongoing process that involves a combination of risk assessment, testing, observation, training, and data analysis. Safety in human-robot collaboration refers to the measures and technologies that are used to ensure the safety of humans who are working with robots. This can include measures to prevent robots from causing physical or cognitive harm to humans, as well as technologies that allow robots to work alongside humans as team member.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Live lab 1.4 is based on IMR, a collaborative robot cell (Intelligent Motion and Robotics) technology. The cell consists of a UR10 robot arm with a gravity compensation controller, allowing it to perform kinaesthetic tasks through human demonstrations, designed in live lab 1.3. The robot is also equipped with a Robotiq 3-Finger Adaptive Robot Gripper with continuous grasp capabilities and a Microsoft Kinect v2 RGB-D camera mounted above it for workspace and object tracking. Lab 1.4 focuses on the safety of the operator in task demonstration as well as the assessment of human performance in kinaesthetic teaching. The project started with the risk assessment the risk assessment paradigm which addressed the use of hybrid Standarization format and focuses on the functional safety of cobot cell. The objective is to make safe and reliable human robot interaction and the focal standard in our application will be ISO 10218 and ISO/TS 15066 standards with other normative standards. The risk assessment is performed using Pilz safety assessment protocols.

As a part of safety assessment in task demonstration, it is necessary to assess the human performance, because human is in the teaching and supervision role. The performance of human robot may affect the overall safety perception and productivity of collaborative cell. This assessment can provide valuable insight into the collaboration's effectiveness and identify areas where the collaboration could be improved. One approach to assessing human performance in a collaborative task with a robot is to use a within-subjects design, where the same participants perform the task both with and without the robot. Live Lab 1.4 aims to conduct an experiment to measure the human performance in kinaesthetic task demonstration. The potential tool to collect human performance data will be physiological sensors and vision system.

## Live-Lab 2: Making the Automotive Factory floor more reliable (IVECO)

Live Lab 2 is focused on manufacturing operations in a large-scale automotive plant. The project will collect data during manufacturing and model the human operator’s performance versus task complexity. The objective is to exploit this data to optimize human performance while simultaneously predicting anomalies and scheduling maintenance events.

Despite increasing automation, the manufacturing sector is still widely based on human operations. Human operators still play a crucial role in many aspects of the industry. This is because manufacturing processes often require a high degree of flexibility and adaptability, which can be difficult for machines to replicate. Additionally, human operators can make judgments and decisions based on their experience and expertise, which can be valuable in ensuring the quality and efficiency of manufacturing processes. As a result, the manufacturing sector will likely continue to rely on human operators for the foreseeable future. The automotive sector, for instance, is based on assembly lines, where the automation process is becoming increasingly complex. However, as automation technology has advanced, assembly lines have become much more complex, with robots and other automated systems performing a wide range of tasks. Different operators contribute to assembly products, which require different operational capabilities and a multi-faceted approach for analysing critical-safety procedures and making technological decisions. IVECO is an automotive company where production systems are based on assembly lines that require the interaction between highly automated workstations and highly trained human resources. The researchers here will test human-in-the–loop-

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

automation performance in the context of different workstations. Human Factors (HF) are a critical component of employers' safety. This is because many workplace accidents and injuries are caused by factors related to human behaviour and decision-making, such as fatigue, stress, and distractions. By understanding and addressing these human factors, employers can take steps to reduce the likelihood of accidents and injuries in the workplace. A practical way to assess human performance modelling, as the reliability of individuals to perform a specific duty, can help identify critical scenarios in manufacturing plans. When an operator has enough capabilities to perform a complex task, the probability of an accident or error is reduced. This is because they are better equipped to handle the challenges of the task and can apply their expertise to effectively complete it. According to the kind of tasks involved in the assembly line, the CISC project will collect data from the operators themselves, such as physiological parameters or numbers of human errors performed during the task execution, to assess the capabilities of the operators to perform the given task in a real-time. While on the other end, accurate sensory data from the machinery interaction (such as welding operations) performance will also be collected (parts completed without errors, unsafe conditions etc.) and environmental conditions. The human performance model will include all the data collected on the shop floor to improve production quality and efficiency and reduce the human errors that could arise in incidents or accidents.

*Live-Lab 2.1. Anomaly prediction and predictive maintenance. ESR 2. Devesh Jawla.*

The CISC project will exploit sensor data for predictive maintenance and fault/anomaly detection. We can use ML techniques to predict faults and/or schedule maintenance work to reduce losses and increase efficiency. If they require a fault detection mechanism, ML techniques will be utilized, and since these are rare in occurrence, active learning could provide an optimal and novel solution. By using Bayesian neural networks, we can also offer credible predictions. Fault detection data could be in the form of human sensor data or IoT sensor data from machinery, which are safety critical.

Reducing the number of assembly line breakdowns and reducing the number of human errors is two critical concerns for the industry. Many sensors collect data from the shop floor in the automotive sector, which can help detect and prevent machine failures and human errors. Machinery downtime and servicing could be reduced considerably if intelligent systems predicted maintenance needs well in time. Similarly, predicting their alertness level can prevent human errors and accidents. These two relevant concerns for the automotive industry can be addressed using ML/AI techniques for anomaly detection.

Anomaly Detection deals with the identification of abnormal data points which deviate significantly from the usual data. Its applications are numerous, and in the context of industry and IoT, it is used to identify safety-critical scenarios, for example, detecting the malfunction of equipment in production assembly or detecting faulty products coming out of an assembly. This is known as predictive maintenance, which monitors equipment condition using machine learning and offers cost savings because maintenance is performed at the warranted time and prevents accidents by warning us before they happen.

*Live-Lab 2.3: Design experiments in an automotive sector line work environment for Human-machine performance monitoring and prediction. ESR 11. Carlos Albarran Morillo*

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

The human performance (HP) modelling can be described as the result of an interaction between the level of skill demanded to achieve a given job in a working place with the capabilities of the employees assigned to it. In particular, in (Leva et al. 2016, Comberti et al. 2018, Leva et al. 2022), it is shown that human performance could be represented as directly dependent on two macro-factors:

- Task Complexity (TC): assessed through Mental Workload (MW) and Physical Workload (PW), both associated to each activity identified and analysed in the assembly line.
- Human Capability (HC): it represents the skills of workers under the actual working states, including the physical, mental and cognitive abilities of each employee.

It is applied this framework to assembly line work, the HC component can be grounded in three quantifiable capabilities: Memory, Manual and Physical skills assessed using the "ability corners". The TC is estimated by assessing the observable variables related to Mental and Physical Workload and expressed in terms of indices harmonized in a Likert scale. The HP model results are addressed using a matching index matrix that compared the required aptitudes for each workstation with the harmonized recorded skillsets of each worker in a static way.

The recent technological advances in various fields, including artificial intelligence (AI), allow enriching the human performance modelling. Since the dawn of the era of modern computers, some tasks, particularly repetitive ones, are best performed by machines. However, nowadays, more than that is demanded. Emerging technologies, such as wearable devices and machine learning algorithms, aim to improve operator safety, performance, and well-being in various settings, including manufacturing, aviation, and healthcare. These technologies can be used to monitor the internal state of operators, such as their physiological signals and cognitive workload. In addition, they can provide real-time feedback to help operators maintain optimal levels of performance and well-being. Additionally, these technologies can be used to enhance the capabilities of the workstation-operator interaction by providing operators with real-time information and assistance and by automating tasks that are repetitive or dangerous. This can help to improve the efficiency and effectiveness of work processes and can also help to reduce the likelihood of accidents and injuries. As a result, operators will have to acquire a broader range of specific skills. They will have more and more often to combine traditional task-associated expertise with computer science one (Albarrán Morillo et al. 2022).

The human performance model is a framework to understand and predict how human operators will behave in a given situation. This model considers factors such as the individual's skills, knowledge, and experience, as well as the characteristics of the task and the work environment. By incorporating this model into safety-critical systems, we can better understand how humans will interact with these systems and identify potential areas where errors or accidents may occur.

We can take several steps to enhance and deploy the human performance model within safety-critical systems. First, we can gather data on human performance in various tasks and environments to better understand the factors that influence how humans behave. This data can be used to develop and refine the human performance model, making it more accurate and predictive.

Next, we can use the human performance model to monitor and evaluate the performance of human operators in safety-critical systems. This can help us identify areas where operator performance may fall below expectations and implement training or other interventions to improve performance. Using the human performance model, we can ensure that safety-critical systems are designed and operated to maximise operator performance and minimise the likelihood of errors or accidents.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

## Live-Lab 3: Assisting Human-decision Making

Live Lab 3 is concerned with addressing alarm handling and process control design and management challenges. Here the ESRs investigate both human and system behaviour during alarm management and human-in-the-loop process control. Therefore, the elements and components of interaction; control room operators, human system interfaces (HSI), decision support systems, are of interest to the researchers. The ESRs through a selected case study would work on their respective thesis goals and as a team to address the Live Lab 3 concerns.

***Case Study***

A case study and scenarios on a simulated formaldehyde production facility has been proposed for experimental studies. The simulator has been developed in the SAfeR group of Politecnico di Torino, Italy (Micaela Demichela, 2017).

Simulated plant: It produces around 10000 kg/h of 30% formaldehyde solution, operating the partial oxidation of methanol with air. For Live Lab 3, there have been variations and optimisation to the referenced simulator (see Figure 4.3.1).

The plant is made of three sections:

- Feed section,
- Heat and recovery section: three heat exchangers
- Reaction and Separation Section.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

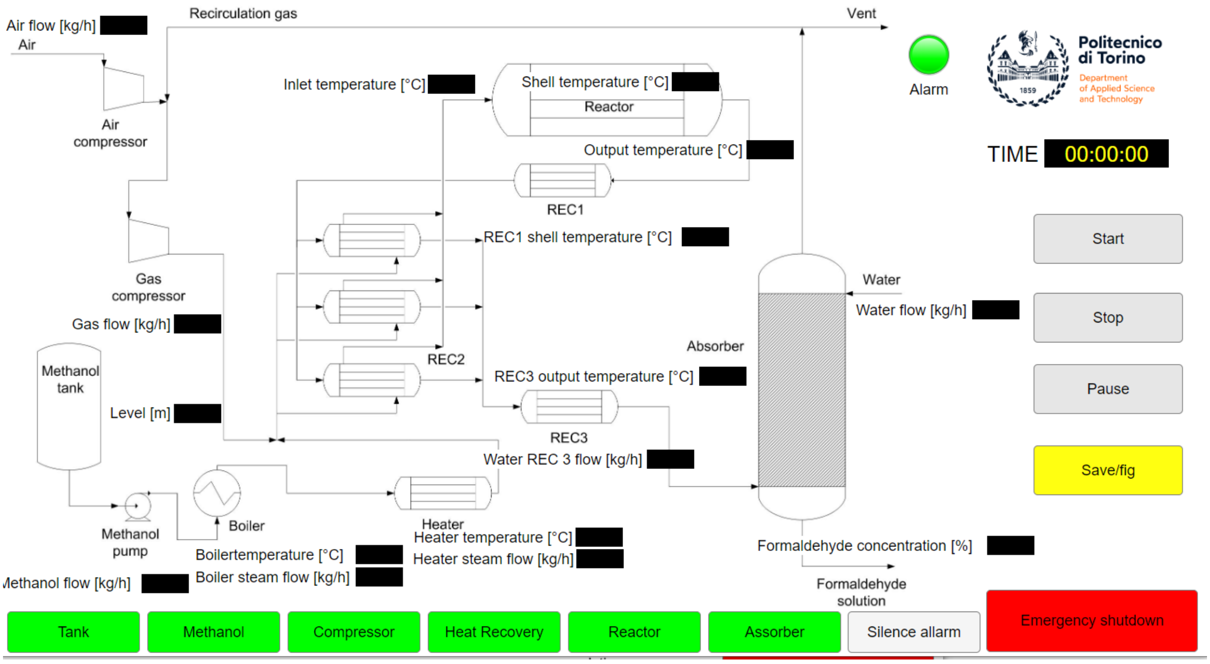



Figure 4.3.1: Interface overview

***Simulations: process safety related Scenarios***

The scenarios are broken down according to increasing complexities:

- Scenario 1: failure of the pressure control logic of the methanol tank
- Scenario 2: failure of the pressure control valve of the tank
- Scenario 3: failure of cooling water temperature control logic

Through this experimental study, the ESRs can monitor, collect data, and analyse both operators' and system behaviour from the human machine interaction under the three scenarios. Also, developed support systems will be tested for their impact on operators' behaviour and performance.

*Live-Lab 3.3. (ESR#4. Chidera Winifred Amazu)* Process safety data modelling for human-in-the-loop configurations in process control

***Process safety***

In safety-critical systems like several process plants, safety is known to be of paramount importance. However, this is compromised by many factors, of which human error is considered the leading contributor. Despite the emphasis on human error, there are still gaps in the way safety analysis is carried out within the process and energy industries. This gap is because of the lack of inclusion of human and organisational factors, both as source of errors, but also for recovery capabilities.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

***Human-in-the-loop configurations***

Recently, considering that operators, especially control room operators, must deal with process and equipment complexity, and increasing automation, a lighter physical involvement of operators is counter-balanced by the increased cognitive load generated by e.g., an increasing number of alarms, especially in abnormal situations and systems failures. A human-centered approach to safety analysis based on human-in-the-loop (HITL) settings would be ideal to face these incoming issues. Also, such an approach should address gaps identified in current efforts to assess human performance though human reliability methods.


Figure 4.3.3.1: Operator-machine/computer interaction loop

***Objective***

The goal of this research is to develop an innovative methodological framework and tools for decision support in process safety critical status (considering technical and HOF data) of HITL configurations and for process control optimisation and process safety.

Therefore, Live-Lab 3.3 is concerned with process safety data modelling of the HITL in process control, that is, modelling both operators' behaviour and system behaviour for an overall process safety analysis. ESR 4 incorporates the following questions in her model; the process safety related event or scenario (that is, what is happening or can go wrong - taking an a priori approach), the components of interaction (that is, what can be seen or will be interacting to address the process safety event, e.g., the human system interfaces, procedures etc.) and finally what the operator does in response. This would be compared subsequently with the system behaviour given the operators' response. Performance shaping factors and key process safety and control indicators for decision-making would be identified in Live Lab 3.3.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

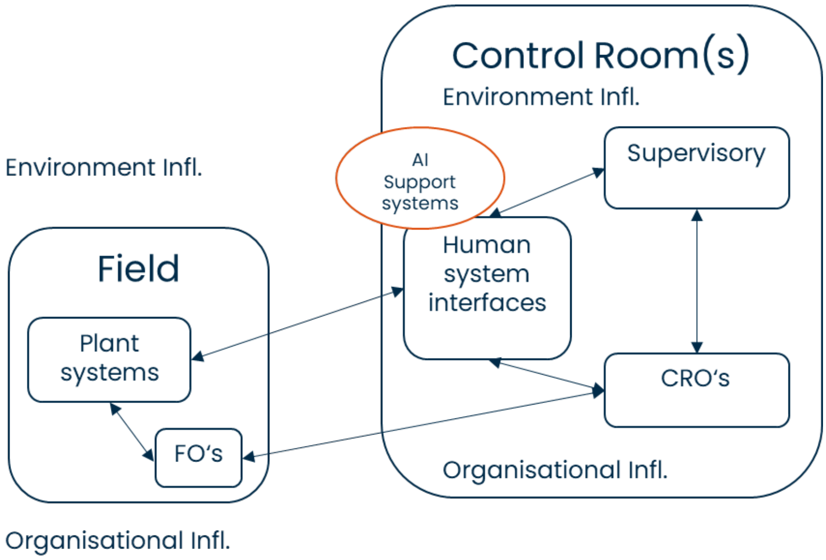



Figure 4.3.3.2: Interacting components

***Type of data***

- Qualitative data: on situational awareness using questionnaires like SAGAT and SWAT
- Qualitative data: on workload using NASA TLX
- Qualitative data: on Trust using trust survey
- Expert analysis: Qualitative heuristic evaluations of interfaces
- Walkthroughs and think aloud with participants
- Psychophysiological data from Eye tracking and heart rate measures
- Audio recording for noise (environment study)
- System behaviour data from log, e.g., on alarms, error rates, interface interactions, process variables.

***Input for model:***

- Tasks and performance shaping factors (PSF) details
- Human error probabilities
- System behaviour from defined metrics
- Consequence analysis

This research work would be carried out using the Live Lab 3 defined case studies.

*Live-Lab 3.4. (ESR#5. Ammar Abbas)* Process control optimization with alarm reduction and prioritization in an online human-in-the-loop setting using deep reinforcement learning.

***Objectives***

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

- Reduce the number of alarms by optimizing the process control through deep reinforcement learning Alarm suppression and prioritization using deep reinforcement learning
- Suggest control actions and/or setpoint to the operator
- Explainability for the proposed actions/ suggestions
- Root cause failure analysis using Bayesian-based methods (such as the hidden Markov model)
- Disturbance rejection and setpoint tracking using deep reinforcement learning
- Residual Policy Learning (RPL) to add a correction factor to the output of the controller response and then feed it to the plant for minimizing alarm flood
- To be able to incorporate human ergonomics as the decision-making factor (input) for reinforcement learning

***Input Data:***

*Process variables:* process variables are the system/environment variable from which the Reinforcement Learning (RL) agent is informed about the state of the system

*Control Variables:* control variables are the variables that may be manipulated by the RL agent to optimize the process based on the information gathered from the system state.

*Process and Control variables thresholds:* thresholds are necessary for the RL agent to know if the system state crosses the normal thresholds for process variables and enters an abnormal situation. For the control variables, the thresholds indicate the bounds for the control for the RL agent.

*Ground Truth:* ground truth data about the process abnormality helps identify the abnormal situation and the system response.

*Expert behavior:* it is not essential, however, can help RL agents to pre-train to avoid random exploration for safety-critical systems.

*Human Performance Measurements*: human performance measurement can help the RL agent to stay aware of the human state and collaborate effectively in a Human-AI collaborative environment. RL agent can automatize process optimization once it identifies a stressful situation.

***Snapshot of a sample data:***

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

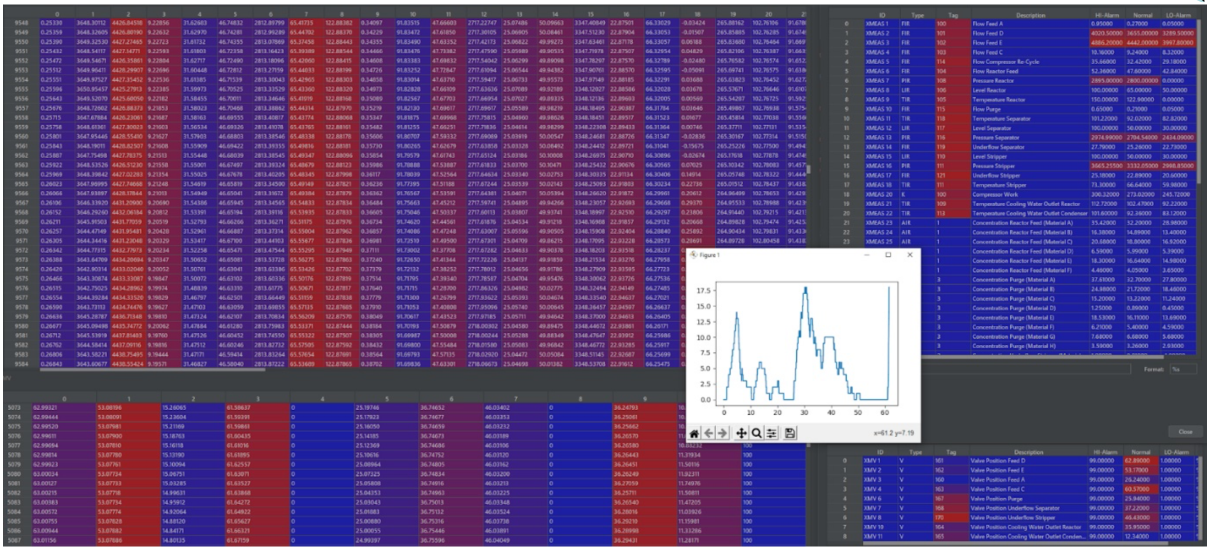


Figure 4.3.4.1: Sample dataset

***Methodology:***

*General Framework:* a general framework is illustrated in Figure 4.3.4.2 for a specialized RL agent which provides interpretations for humans along with the RL actions through a probabilistic model-based RL. The RL agent specializes in a specific situation and provides a hierarchical architecture (Abbas, Interpretable Input-Output Hidden Markov Model-Based Deep Reinforcement Learning for the Predictive Maintenance of Turbofan Engines, 2022).

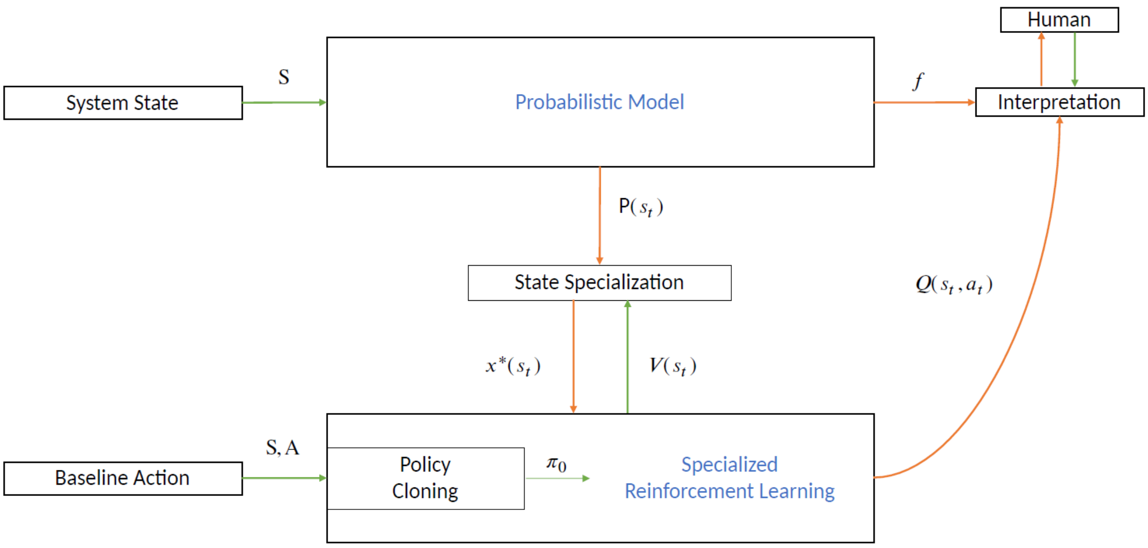



Figure 4.3.4.2: Specialized Reinforcement Learning Agent (SRLA) (Abbas, Interpretable Input-Output Hidden Markov Model-Based Deep Reinforcement Learning for the Predictive Maintenance of Turbofan Engines, 2022).

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

*Case Study (Process Control Formaldehyde Plant):* the general framework is specified and used for the process control.

*Schema:* the overall schematic diagram of the RL integration with process optimization (Abbas, Deep Residual Policy Reinforcement Learning as a Corrective Term in Process Control for Alarm Reduction: A Preliminary Report., 2022) is provided in Figure 4.3.4.3.

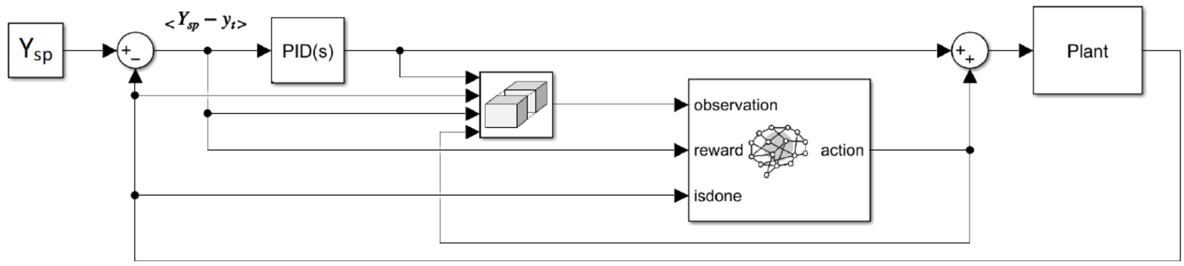



Figure 4.3.4.3: Schematic diagram of integration of RL to process control (Abbas, Deep Residual Policy Reinforcement Learning as a Corrective Term in Process Control for Alarm Reduction: A Preliminary Report., 2022).

*Framework:* the specific framework used for the case study is shown in Figure 4.3.4.4. Input-Output Hidden Markov Model (IOHMM) is used as the probabilistic model and an Actor-Critic architecture is used for the RL agent.

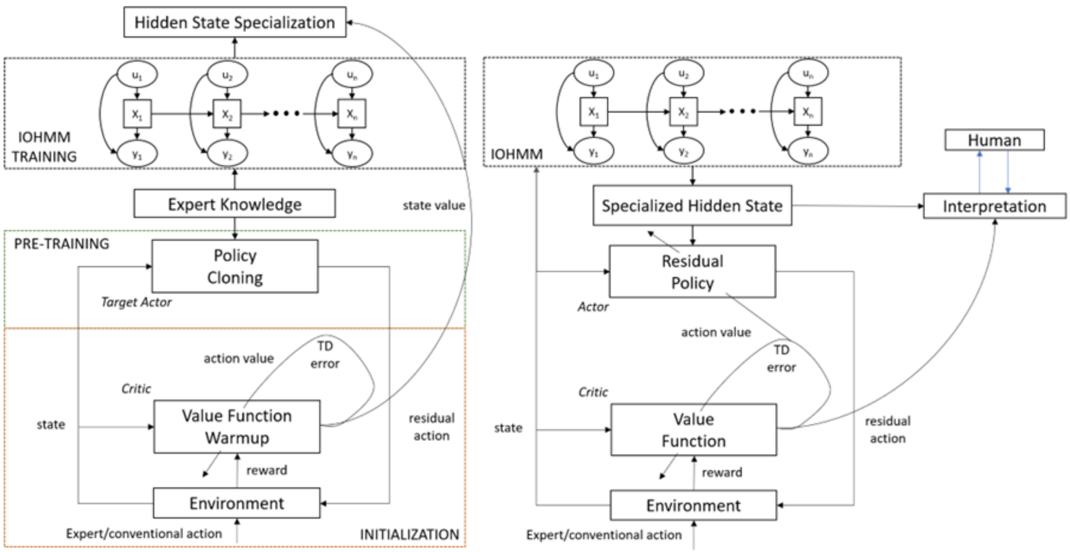



Figure 4.3.4.4: (Left) IOHMM training with classification through the value function and behavioral cloning of the base policy followed by the initialization of the critic (Right) Online training and fine-tuning phase of the framework activated on the specialized hidden state

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

*Live-Lab 3.5. (ESR#6 Milos Pusica)* Designed an experiment for evaluation of impact of simple/complicated visual assembly line instructions on operator's mental workload.

Ergonomics is a very important aspect of a workplace from the point of view of productivity, workload, fatigue and overall well-being of an employee. It is even more important in the cases of monotonous jobs, like assembly line jobs, when employees spend vast majority of their working hours in the same position. One of the factors that have an impact on neuroergonomics in assembly line workplaces is the complexity of the instructions given to a worker.

worker's performance and workload. The experiment was conducted in laboratory conditions in modular and adaptive laboratory set-up for neuroergonomics and human-robot interaction research (Savković M, 2022). A worker was performing manual assembly task on small plexiglass plates, given multiple schemes images on a screen, and asked to assemble empty physical plates in the same way as shown on images (Figure 4.3.5.1). Schemes images were of two types - low complexity and high complexity schemes (Figure 4.3.5.2).

*Live-Lab 3.6 (ESR#9 Doaa Almhaithawi)* Using latent spaces to explore security challenging such as intrusion detection and pattern recognition

As AI has proven itself as a helpful tool for decision making, the research is still active in finding more accurate, real-time, and independent methodologies in this regard. On the other hand, latent spaces and representation is an extremely attractive domain for researchers due to his state-of-the-art results in many vital fields, the main idea behind these spaces are the study and exploration of the data structure which will be helpful in any task or application, ex: anomaly detection, pattern recognition, data augmentation. However, the starting point should be the building of these spaces as an accurate representation of the corresponding data. Therefore, our contribution to the human-decision applications is highly dependent on the data itself, so far, we are in the theoretical phase where we are studying the techniques and methodologies used in this field, for applying them later on each available task.

*Live-Lab 3.8. (ESR#14 Andres Alonso Perez)* Testing of different ML approaches for EEG feature extraction to detect operator's mental workload

In working environments in which the Human actively collaborates with a machine (e.g., in the industry), there is a critical need to properly define the information shared the information between the machine and the worker, to organize the tasks that the human performs, and to assert the mental state of the human during the tasks.

That would allow not only a better performance and optimization of the overall process, but a decrease of the chances of Human Errors that can affect the process, an improvement in the Human Safety in the working environment, and a reduction in the workload associated with the tasks developed (with the consequent improvement of the working conditions.

According to (Coster, 2017), there is a space in which the distribution of the tasks can be performed if the levels of Workload and Situational Awareness are balanced. That leads to the question: what is an

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

effective way to assess Mental Workload and Situational Awareness in a working environment so an efficient task allocation can be performed?

There are many sources on the neuroergonomics field that discuss how the physiological information from the worker obtained by certain devices (e.g., EEG), can be used to assess their mental state and their chances of making an error or having their performance decreased (Dehais, Lafont, Roy, & Fairclough, 2020).

About Situational Awareness, it is common consensus that it can only be obtained with either subjective evaluation from the worker or from external evaluation from an expert. Focusing on mental workload, there is a trend to use EEG, with different features extracted from it as indicators of Mental Workload (e.g., theta frontal power divided by alpha parietal power (Holm, Lukander, Korpela, Sallinen, & Müller, 2009)). Other studies have also focused on the possibility that there is a feature able to represent Mental Workload and can be transferable between tasks (Yufeng, et al., 2021).

Nevertheless, some recent studies have proven Deep Learning to be able to extract more significant information from EEG, if compared to the handcrafted features. Some other studies, both with handcrafted measures and with the application of Deep Learning, suggest that using the relation between the electrodes of EEG in the form of a graph can result in creating significant features from EEG.

# 3. Data integration need for Artificial Intelligence in Collaborative Intelligence

For the general description of LiveLabs, please refer to Deliverable 5.1.( D5.1_Ready.docx). A description summary is provided in Section 2.

## Live-Lab 1.1. Human Ergonomics. ESR 7. Carlo Caiazzo.

The subject is connected to the design of the overall architecture of the system that will be necessary to use record and interpret physiological data for physical ergonomic assessment in Industry 4.0 settings for everyday tasks. The proposed modular workstation in the Live-Lab gives the go-ahead to perform neuroergonomic experiments in manufacturing scenarios, involving state-of-the-art systems and sensors supporting the operator to accomplish the tasks. During the experiments, muscle and brain activity are monitored in real time during the performance of repetitive, monotonous assembly activities by EEG and EMG sensors. The collection of these data is the input to analyse the physiological performance of the operator during the tasks. An innovative EEG system was used to design and conduct the neuroergonomics experiment. The SMARTING wireless EEG system (mBrainTrain, Serbia) was used for EEG signal acquisition. This device has the ability to record EEG signals with a sampling frequency of 500 Hz and a 24-bit data resolution. The SMARTING EEG amplifier (85x51x12mm, 60g) was connected to a 24-channel EEG cap in the occipital region of the head using an elastic band. The connection between the amplifier and the computer was made using a Bluetooth connection. For EMG measurements, the muscleBAN (PLUX Wireless Biosignals, Portugal) was used. This wearable wireless (Bluetooth or Bluetooth Low Energy data transmission) device combines a single-channel

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

EMG sensor, triaxial accelerometer and magnetometer and in that way enables real-time acquisition with up to 16-bit resolution at up to a 1000Hz sampling rate.


Fig.3.1.1.1: EEG cap and EMG sensors

Electroencephalography (EEG)is one of the most common methods for assessing the cognitive state of the operator. Bakshi (2018) detected the cognitive workload of 28 subjects through EEG. Moreover, several authors conducted neuroergonomics tests of brain activity during manual assembly of a hose.Other studies identified the relationship between cognitive load and changes in the EEG signal (Antonenko et al., 2010; Brouwer et al., 2015; Charles and Nixon, 2019).

EEG signals are directly correlated with mental demand experienced during the task (Brookings et al., 1996). The captured EEG signals are analyzed to identify their features by fusing them to define the overall brain activity. They enable direct measurement of brain activity in real time. Moreover, they can estimate the quantitative assessment of alertness levels, which requires expensive computational signal processing. The main advantage of EEG is the possibility of objective measurement (as opposed to subjective methods of self-assessment) of workers' attention in real time (Mijović et al., 2016).

In industrial scenarios, EEG is widely used to assess the cognitive state and mental workload of workers (Infantolino and Miller, 2014). Foong et al. (2019) used EEG to identify drowsiness of 29 subjects. Numerous studies evaluated the measurement of cognitive load via the EEG signal (Scalera et al., 2020).

On the other hand, EMG is the most popular and commonly used method for detecting the occurrence and development of muscle fatigue (Freitas et al., 2019). EMG signals represent neuromuscular activities of the human body. They are used to monitor workers' muscle condition and find the maximum lifting load, lifting height, and number of repetitions that the workers are able to handle before experiencing fatigue, all for the purpose of avoiding overexertion (Gevins et al., 1995).

EMG ergonomics applications are the most widely and successfully used in industry for real-time fatigue monitoring, musculoskeletal risk assessment, and assisted handling devices. EMG is the most commonly used tool in many research papers (He, Zhu, 2017). In contrast to the subjective methods of measuring muscle activity, EMG is characterized by objectivity and reliability. The study of EMG signals can help assess functions at the muscle level and at the level of the nervous system,which controls the muscles.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Bosch et al. (2007) showed EMG manifestations of muscle fatigue of the trapezius muscles during normal (8-hour) and extended (9.5-hour) working days involving light manual work. Bennie et al. (2002) also simulated 8h-hour working days using EMG measurements.

The main goal of monitoring muscle activity in the neck is to determine the load and strain of the neck muscles in order to examine the frequency of neck pain and the onset of the first symptoms of MSDs. Muscle activity was monitored by placing EMG sensors on the trapezius muscles on the subject's neck on the left and right sides (Savković et al., 2022). Brain activity was monitored in order to examine the subject's mental fatigue and, on the basis of the obtained data, determine when attention and concentration decrease (Savković et al., 2022). An EEG cap with electrodes was placed on the subject's head in order to monitor the brain activity.

## Live-Lab 1.2. Human performance monitoring for human-in-the-loop telerobot operations. ESR 8. Ines Ramos.

The project's goal is to perform human performance-related state prediction in complex human-computer/machine interactions, based on explainable deep learning and multi-modal data fusion techniques. The assessment and data collection will be conducted in a specific collaborative intelligence use-case, Live Lab 1.2, but the developed framework, ideally, can be used in other human-computer interaction contexts.

Human performance will be modelled by the neuroergonomics framework developed by (Dehais et al. 2020), that maps human performance degraded states, such as mind wandering, effort withdrawal, perseveration, inattentional blindness and deafness to two dimensions, arousal and task engagement. The framework proposes an alternative human performance analysis to the commonly used concept of mental workload, associated with limited information processing resources, that does not account for all of the identified performance degraded mental states. In this framework, the concept of task engagement captures the goal-oriented aspect of cognition and arousal captures levels of disengagement that happen in states of high arousal or low arousal. These two dimensions will be used to identify different operator internal states that are induced while performing a teleoperation task by variations of the HMI's features. Indicator values of arousal and tasks engagement will be measured, and a clustering analysis will be performed to determine number of elicited states and experimental conditions associated with it. These values and performance measures will be used to label human performance-related states (ground-truth), to be used as a dataset for deep learning model training.

For monitoring of arousal and task engagement in human-computer/machine interactions, two types of physiological/bio sensors stand out, EEG and Eye-tracking. EEG provides different measures of neural activity for different mental states/mental constructs, such as mental workload (Ren et al. 2019)(Yang et al. 2019) (P. Zhang et al. 2019) (Meteier et al. 2021), fatigue (Ahn et al. 2016) (Dehais et al. 2019), situational awareness (T. Zhang et al. 2020), task engagement (Dehais et al. 2019), attention (Mijović et al. 2017), and affective state (Song et al. 2018) (Zhong et al. 2019) (Yin et al. 2021), with a high temporal resolution and low latency of the changes that occur in response to external stimuli or internal mental processes. Despite the potential of this modality, compared to other cheaper and more convenient sensors to set-up and use (such as ECG, respiration belts, plethysmography or skin

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

conductance sensors), it has received less attention in practical applications. However, with the development of mobile EEG technology, this type of physiological sensors is moving closer to widespread use in a variety of workspaces. Eye-tracking can be performed with less intrusive sensors (e.g. glasses or sensors attached to a screen) and provide temporal and spatial information about blinks, gaze fixations, saccades and pupil changes, albeit with much lower resolution compared to EEG. Furthermore, it can provide visual information of where the user is looking at in the environment/interface/screen. These features can be utilized as indicators of attentional states, visual task engagement, fatigue, mental workload and other mental constructs, however, for emotional arousal level in specific the pupil diameter has been a known proxy that changes with the sympathetic autonomic activity of the nervous system (Bradley et al. 2008).

For the Live Lab 1.2 a mobile EEG cap from mBrainTrain (Smarting mobi, mBrainTrain, Serbia) be used with a sampling rate range between 250Hz-500Hz and a Tobii Pro Nano eye-tracker mounted on a monitor, with sampling rate of 60Hz. Beyond using the eye-tracker to understand gaze patterns in the use of the telerobot visual user interface (Figure 3.1.2.1 image on the bottom right) a multi-modal approach using both modalities will be employed to train the deep learning model to predict human performance states (Figure 3.1.2.1 image on the top left). (Poria et al. 2017) defended and demonstrated that a multimodal approach is feasible with the existing state-of-the-art information fusion methods and can lead to performance improvements for affective computing systems. As humans can convey their internal state in a variety of ways (either conscious or unconscious changes in the body and behaviour) that can differ according to the context, analysing human physiological information from multiple sensor modalities combined can capture a more comprehensive picture of the user state and lead to a better performance and reliability in recognizing it. Additionally, the integration of multiple sensors can increase the robustness to noisy data, motion artifacts, data loss due to blinks or head movements or other perturbations to the recording of data, and to the intra and inter-subject variability.

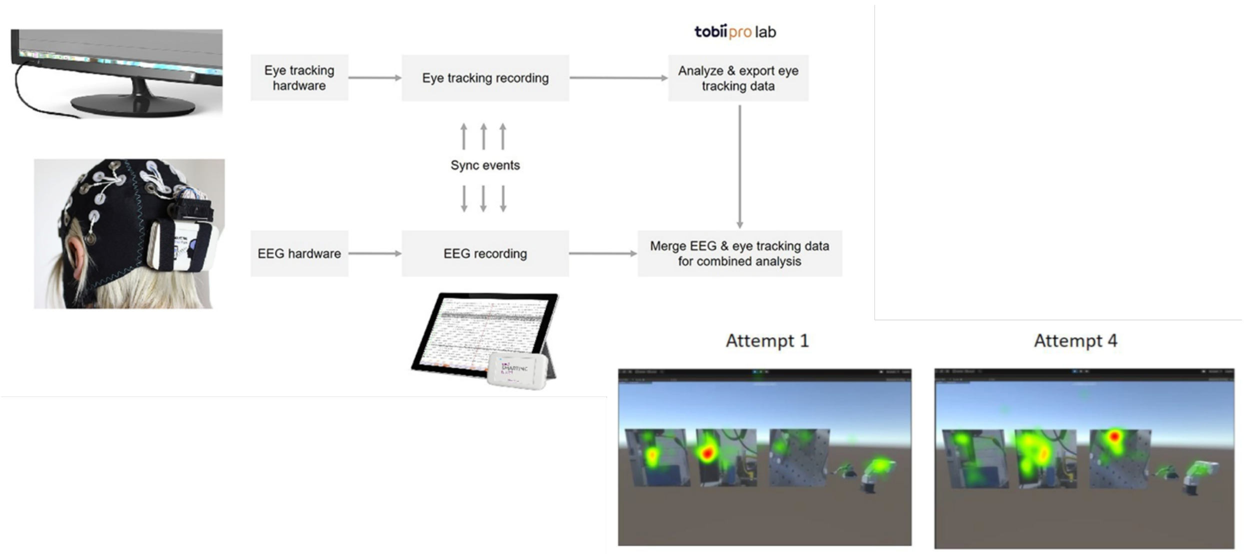



Figure 3.1.2.1 – Schema of the multi-modal data integration and processing pipeline for the Live Lab 1.2 (image on the top left). Visual and haptic user interface of the telebot system. Eye gaze heat map overlapped with the telerobot visual user interface (aggregated values from the entire task duration), obtained with Tobii Pro Lab software.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Data or information fusion techniques have been extensively used for affective state recognition, using mostly audio and visual data for joint speech and facial expression emotion recognition, but there are already many works that have used multi-modal physiological data instead (Healey and Picard 2005) (Putze and Schultz 2014) (Yu et al. 2020) (Wu et al. 2021) (Vortmann, Ceh, and Putze 2022) (Dessai and Virani 2022). The two main types of data fusion are feature-level (or early fusion), that combines multimodality features in a single general feature vector after time synchronization and format alignment, and decision-level (or late fusion), in which the features of different modalities are analysed independently and only at the decision point the results are combined. According to the review of (Shoumy et al. 2020), decision-level fusion has demonstrated to have many benefits compared to feature-level fusion: it is easier to deal with the results of the analysis of individual modalities as they usually have the same representation, while features of different modalities usually have different representations and dimensionality. With late-fusion it is also possible to select which modalities to fuse. Other less common fusion strategies that have been employed in research are: hybrid-fusion, that combines both early and late-fusion techniques; and model-level fusion, that fuse different modalities based on the correlation between the data, the research needs and problem space, using either fusion based on rules (rule-based fusion), fusion based on a classification or fusion based on estimation using Kalman filter.

In this project, the selected data fusion technique and the deep learning model architecture were chosen considering the modalities that are going to be used, the potential applications of the developed and trained model and the need for a level of explainability that is required for safety-critical systems. Accordingly, the first proposal is to develop an end-to-end training framework using a hybrid CNN-LSTM (Convolutional Neural Network- Long Short-Term Memory network) model where the CNN will be used for spatial feature extraction and the LSTM for time dependencies modelling. These models were chosen not only because they are deep neural network architectures specialized in the type of data that will be used, but also to take advantage of the high number of model-specific explainability post-hoc techniques that have been already developed due to the popularity of these models. In particular, we are interested in the self-attention mechanism that can be applied to both models and that can provide some level of interpretability by knowing the times and spatial features that the model attends to when discriminating between performance states. Additionally, the computed attention coefficients can be used for attention-based data fusion of streams with different temporal dynamics and resolutions (as the case of EEG and eye-tracking signals) and provide multi-modal temporal attention.

Furthermore, to exploit the common information between the two physiological modalities about the internal human state, to better disregard other irrelevant information encoded in the signals and to potentially achieve better performance when training with small datasets, a self-supervised learning (SSL) approach using cross-modal clustering could be applied. This type of SSL allows to train the model with a pretext task using pseudo-labels generated by deep unsupervised clustering. The key step to leverage the correlation between these modalities is to perform cross-modal prediction of the clusters using the pseudo-labels generated by the other modality. In this manner the model can better learn the important information and features to extract and generate a better initialization of the model for further fine-tunning with the collected data and classification task of interest. This approach has been applied successfully to the SSL of action recognition using video and audio modalities (Alwassel et al. 2019) and movement classification/decoding using EEG and other kinematic and physiological modalities (Peterson, Rao, and Brunton 2022).

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Other public datasets containing EEG and eye-tracking data (Ceh et al. 2019) (Wei Liu et al. 2021) can also be used in the pretext task and contribute to the generation of more robust and generalizable feature extraction from the SSL. To first understand if this technique can be effectively applied to the modalities of eye-tracking signals and EEG (considering the difference in the temporal resolution), two public datasets with EEG and eye-tracking data for affective recognition will be used for the SSL task with cross-modal deep clustering, to compare the achieved performance when doing no pre-training of the model, when pre-training with within-modality clustering SSL and when pre-training with cross-modality clustering SSL. Additionally, we can study the performance impact of pre-training with another related dataset or with the same dataset used for the fine-tunning supervised task.

A potential application of such model, if demonstrated to have the required robustness, can be monitoring in real-time the internal performance-related state of an operator/user in safety-critical human-machine/computer interactions (e.g. using telerobotics for rescue missions, or other types of systems) and perform performance-driven interface adaptation actions to change this state to a more desirable one or to simply inform the user and give them the decision-making control. For the medical device manufacturing use-case, the probability of serious accidents or failures is very low and the associated risk as well. Nonetheless, such a system could be employed to reduce the probability of damaging equipment to virtually zero and to increase the performance in such a human-in-the-loop task. Similarly, the developed model can also be used in the design stage of the HMI or for evaluation of added intelligent/autonomic features to the system. The real-time continuous monitoring of physiological signals and classification into the states of interest has proved to be an efficient framework for human-centred interface design (Lim et al. 2018), especially compared to the traditional user studies that require explicit feedback and cannot access it in real-time.

Some of the remaining challenges of data integration and processing in this project and Live Lab include dealing the synchronization of the multiple streams, ensuring a non-biased pre-processing of EEG data (can require manual removal of noisy data segments or artifacts), dealing with missing data in one or two modalities, dealing with small datasets and in the case of when applying SSL, dealing with the processing and understanding of data with different formats and schemas. The Tobii Pro Nano eye-tracking device stream and MBrainTrain's Smarting Streamer EEG stream can be temporally synchronized using a Lab Streaming Layer (LSL) or keyboard events specifically set for important events (e.g. start and end of task), although the second option can lead to small time delays in the recorded markers compared to the keyboard events and requires post-processing to align the streams according to the marked events. The pre-processing pipeline of EEG data will be semi-automated and the specific steps, decisions and criteria will be defined according to the existing literature and documented to ensure a transparent and strict pre-processing guideline is followed for all the data samples. A statistical analysis of experiments with humans using the telebot cell will be performed in a pilot study to analyse and validate the impact of the experimental conditions on the physiological signals and task performance, and better understand the temporal dynamics of the changes observed. This will not only provide important knowledge for the design of the deep learning model, but it will also ensure the experiments for data collection will generate meaningful and quality data to train the model. Lastly, the data formats of the generated dataset will be defined to match the most used formats in this research domain (to comply with transferability requirements), which may require data format transformations.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

## Live-Lab 1.3. Programming from demonstration for fast robot programming and human assistance during anomaly detection. ESR 12. Aayush Jain.

Currently, programming a robot constitutes between 35-40% of the cost of deployment in an industry which is 4 to 5 times the cost of the robot itself (Statista Search Department, 2015). As the industrial environments move towards more complex and unstructured settings, the cost of programming will further rise. Additionally, it will be impossible to manually pre-program all the desired robot behaviors in the future. However, it's much easier and natural for a system integrator to demonstrate the desired behavior like how we teach humans new tasks by showing (Osa. T. et. al., 2018). Therefore, programming by demonstration is a promising approach to teaching robots on the go.

As the aim of PbD is to learn/abstract the desired robot behavior policy π from a dataset D, PbD is closely related to the field of structured prediction (Daumé III et al., 2009, Ratliff et al.,2006, Taskar, 2005) from supervised learning. The task here is to learn a mapping from inputs x (dataset) to a complex but structured output y (task and motion plans). Moreover, we require one-shot learning techniques as it is expected the ITL system immediately learns from each interaction with the expert (J. E. Laird et. al. 2017). Therefore, ITL primarily uses learning techniques other than deep learning and reinforcement learning which generally require a large dataset to derive behaviors. As G. Lentini et. al., 2020, pointed out, a combination of data-based learning (eg. Artificial Neural Networks) and logic-based programming (Good Old-Fashioned Artificial Intelligence) could be a promising approach for ITL.

Machine learning methods have had huge success in a wide range of robotic applications over the past decade and PbD techniques leverage many of these standard methods. In any learning process, data collection is critical to a successful learning process. The quality and performance of learning are directly influenced by the size and variety of the training and testing data. The larger and more diverse the data, the faster and more accurate the system will be able to learn, and the better it will be at generalizing to new situations. However, it is nearly impossible to collect data for all possible situations. Furthermore, learning from humans poses additional challenges like correspondence problems, limited human patience, and user's teaching capabilities.

A typical PbD pipeline consists of a human-machine interface to record demonstrations by experts which are further used to learn a generalized policy to reproduce the demonstrated behavior from the recorded data (Chernova. S. et. al., 2014). Our proposed ITL framework consists of four modules- the human-robot interaction (HRI) module, the learning module, the execution module, and the perception module, which are highlighted with different colors in Figure a. Also, the ITL's workflow is organized into three phases: offline learning, execution, and online learning represented by different arrows in Figure a.

The HRI module enables the operator to record the dataset via kinesthetic and visual task demonstrations. In addition, there is a human supervision interface where an operator monitors the system's decision-making process and real-time task execution information through a BT. In the learning module, information at the task, motion, and state-space levels are abstracted to generate a reactive task policy in the form of a BT. It also refines the policy using data from unsuccessful and successful task attempts.

The execution module is integrated with the automatically generated BT to adapt the high-level and low-level policy of the system depending on the real-time surrounding conditions. The adapted

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

motion trajectory is then transferred to the hardware. Lastly, the perception module is equipped with proprioceptive and exteroceptive sensors, through which the demonstration data are collected and real-time adaptation is achieved. Furthermore, it also monitors the task execution and prompts the operator if a failure is detected in the BT with information about the failure and the subtree where it failed.

In the offline learning phase, a task demonstration is recorded in the form of multivariate-time series data. Elthough each demonstration might look like a single behavior, however, it may consist of a sequence of different primitive actions (Osa. T. et. al., 2018). Therefore, the demonstrated task is then decomposed/segmented into a sequence of sub-tasks or actions. Each action is described in the form of a tuple, similar to the PDDL description- <action, pre-condition, effect> which is logically abstracted from the demonstrated data (S. Jimenez et. al. 2012). Actions are encoded using dynamic movement primitives (DMP) (A. J. Ispeert et. al., 2013). These action descriptions are used to generate BT (M. Colledanchise et. al., 2021) based on a set of pre-defined rules.

The execution phase is responsible for the implementation of the generated BT. The high-level objective of “what task to perform?" calls the previously learned actions and sequences that are encoded in the hierarchy of BT. The low-level objective “how to perform these actions?" is adapted by modifying the learned action (DMP) based on the current object pose and target pose. Finally, the computed trajectories of each action are executed in the learned sequence to perform the task.

Although a single demonstration is sufficient to abstract the necessary information about the demonstrated task, the dataset may not cover all possible situations that a robot may encounter during the execution as it is too expensive and time-consuming to collect all this data (J. E. Laird et. al., 2017). Also, the ever-changing dynamics of the environment and sensor errors may cause failures in execution (S. Niekum et. al. 2013). Thus, a way to incrementally improve the task structure from failures is necessary. In such situations, operator intervention, where feedback in the form of corrective demonstrations, could prove to be helpful. Furthermore, learning from successful executions could be used to refine the skill model.

A failure detection module that is trained using the data from successful task execution runs in parallel to isolate and classify the potential failure conditions during execution. At the time of failure, the operator has the option to intervene to provide a corrective demonstration for failure recovery, if the skill model is insufficient to handle the changes. Using this new demonstration, we can generate an extra branch in the sub-tree where the failure occurred, which handles the recovery behaviour. Through this proposed method we will be able to learn a reactive task policy from a single demonstration and incrementally improve the task structure by learning from failed execution and human intervention.

## Live-Lab 1.4. Introducing AI based safety function using multimodal data pipeline and assessment of safety function under safety integrity levels (SIL) and IEC 61508. ESR 13. Shakra Mehak.

AI-based safety functions in human-robot collaboration refer to the use of artificial intelligence (AI) technologies to enhance the safety of human-robot interactions. These safety functions can include

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

the use of machine learning algorithms to detect potential hazards or collisions, and to adjust the robot's behaviour accordingly to avoid harm to the human. This can help to prevent accidents and injuries, and to make human-robot collaboration safer and more efficient. To design our safety function, we proposed operator's action recognition as a potential tool.

Broadly, techniques for human action recognition comprise three steps: feature extraction, attribute selection, and action classification. A survey of available techniques for human action recognition based on depth data is provided in(Wu et al., 2017). However, Industrial applications need to have a quick reaction time to ensure cooperation and safety. Most techniques presume that actions are temporally segmented (Khan et al., 2022), whereas, in real-world circumstances, actions occur in conjunction with other actions. While assuming the availability of segmented actions, the method described in (Doshi & Yilmaz, 2022) also specifies that only one action is occurring at any given moment; for instance, if a person is walking and waving his hands, only one of the actions would be classified. For industrial human-robot interaction, numerous significant actions coincide. In addition, effective communication and coordination can improve collaboration in safety-critical scenarios. In this regard, operator's action recognition modelling incorporates these characteristics in robot agents. Initially we designed a function using deep learning model. The model is trained on various RGB videos to identify various industrial low-level actions. Potentially data collected from live lab 1.4 will be used to train model.

In contrast, implementing AI-based safety functions is challenging to verify AI-based safety features; the study will investigate how they might be validated using current industry standards, such as SIL or IEC 61508.

## Live-Lab 1.5. Ethical and Legal Implications of the Use of Real-Time Data-Gathering Devices in the Workplace. ESR 10. Naira Lopez Cañellas

The European Commission's High-Level Expert Group on Artificial Intelligence (HLEGAI) recognises three crucial requirements that AI applications must fulfil: they must be ethical, lawful and robust, both socially and technically (HLEGAI, 2019). HLEGAI's main publication, the Ethics Guidelines for Trustworthy AI, offers a comprehensive framework to analyse the collaborative intelligence setting of Live Lab 1, and evaluate its methods for monitoring the muscle and brain activity of an operator working from a remote location using EEG and EMG sensors.

From the ethical perspective, HLEGAI is concerned with upholding the principles of human dignity and autonomy; harm prevention, and fairness and explicability (idem, 13). Such goals require actively considering the asymmetries of power and vulnerability between different groups. These are found, for the case in hand, between employers and employees, as well as between employees of different demographics, due to their differences in gender, age, ability, socioeconomic background, tenure in the company, etc (idem, 12).

Such differences within the workforce must also be taken into account as part of the duty of technical safety and robustness (ibidem). For instance, any AI-based devices must be tested on, and if necessary, adapted to these different demographics, so that they can be accessed according to each worker's needs in an individualised manner. In the case of Live Lab 1, specifically, workstations

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

must be equivalently equipped in a way that doesn't reinforce their different abilities but enhances them. Whenever necessary, safeguards must be put in place to compensate for possible discriminations which may arise from the deployment of said data-gathering devices. In this regard, it is important to make sure that the values that are expected as output aren't generalised from a standardised sample, but accommodated to employees' different traits.

Ensuring the fulfilment of these conditions requires that the agency and duty of oversight remain in the hands of human agents, so that any potential infringement of fundamental rights can be detected pre-emptively and brought to a resolution according to the standing jurisdiction. The endeavour of putting the human in the centre stretches far beyond fundamental rights, as it is closely linked to democratic values and environmental and societal well-being, as explored further below (idem, 19).

While the Guidelines drafted by the HLEGAI's tackle concerns related to privacy and data governance, these are best dealt with from the legal point of view, and most remarkably, through the EU's General Data Protection Regulation (GDPR). This piece of legislation, adopted in 2016, is concerned with protecting European citizens' personal data, and lays out the obligations of all businesses operating in Europe when it comes to handling the data they collect within the delimitations of the Union (European Commission, 2016).

The rationale for GDPR is to clearly lay out which data can be collected, how it can be used and how it can be shared. Similarly, it seeks to compile the rights of Europeans vis-à-vis the information collected about them, which stretch from the need for explicit consent every time a business requests their data, to the right be forgotten (European Commission, 2016).

One of the areas of potential clash between the regulation and the kind of collaborative intelligence setting present in Live Lab 1 regards GDPR's recognition of the principles of purpose limitation, data minimisation and storage limitation, which contrast the Live Lab's potential for both extensive and intensive data collection. GDPR seeks to embed them in the European data culture in order to diminish the chances of misuse of data as much as possible, while still respecting business' commercial interests. Beyond the restrictions regarding how much data can be collected, the objectives for which it can be used, and for how long it can be held, GDPR demands that data holders make sure their data is accurate, and that its integrity and confidentiality are maintained, including all necessary investments in reliable software and hardware as a precondition to collect such data (idem).

Another of the main concerns in collaborative intelligence workplaces regards data transfers. As long as the data of European citizens is being processed, even if it takes place through third parties located outside of the EU, GDPR must be followed. Further regulations will be introduced under the Data Act and the Data Governance Act, and a myriad of bi- or multilateral agreements on this matter are either underway or promise to be in the near future, as is the case of the EU-US Trade and Technology Council. Until these negotiation processes are concluded, and they are accommodated into the pre-existing architecture of both the EU and its member states (including GDPR and its implementation into each national context), transnational movements of data remain in a fluctuating legal space.

To complete this picture, the proposal for an Artificial Intelligence Liability Directive (AILD) comes into play with the aim of tackling the need for accountability in case of malpractice or misuse of AI

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

applications. Accountability is one of the central values in virtually all AI-related debates (HELGAI, 2019), thus this signifies a crucial step in bridging ethics and regulation in the field of AI.

Until the regulatory environment stabilises, corporations responsible for the heavy collection, use or transfer of data can seek to stay on the safe side by adopting a precautionary approach (European Commission, 2000), for instance by storing the data in the local servers of the location where they were gathered, establishing the highest standards of privacy, safety and security on the market, and steering clear from sharing it with third parties without the explicit consent of the individuals it directly or indirectly relates to. Similarly, they should avoid pooling it together or linking it with any other data that renders the data traceable to the individual from whom it was collected. Given the technical advances in the field that increasingly facilitate de-anonymisation processes, they should seek to invest in and prioritise the protection of personal data – as well as non-personal data that could be harnessed for identification purposes too –, and as a last resource, establish the appropriate liability measures for said attempts at breaching privacy. Somewhat counterintuitively, and depending on the data in question, aggregation can actually help re-identify a sample that was previously anonymised (Murdoch, 2021) – hence the importance of treating data protection on a case-by-case basis, and putting in place tailored strategies for data management.

While the incorporation of cobots in an assembly line often carries with it the promise of a reduction in the quantity, length, difficulty, danger or repetitiveness of the operators' tasks, it should not be assumed as a given. Similarly, it is to be expected that different objectives may clash or impact each other; for instance, the deployment of cobots in order to increase productivity and reduce machine downtime can bring a reduction in workforce numbers and/or a shift in work tasks which may not have been anticipated or planned for. Along the same line, a change in the cobot's learning method in order to diminish its training time can imply taking a shortcut in its deployment, or avoiding the operators' involvement in the discussion, design and implementation of its training program. The search for efficiency should not come at the expense of a deliberative process in which the operators who will be in charge, or bear the burden of the shift, take an active role in the discussion, both directly and through their respective unions and workplace representatives. Most importantly, it is crucial that there are safe, anonymous and responsive avenues for the expression of dissent and to request redress in the face of perceived harm, injustice or wrongdoing. It is crucial to ensure these mechanisms rebalance the traditionally subordinate relationship between employer and employee, so that any questions or issues that may arise can be expressed at the beginning of the cycle (e.g. in the context of Live Labs and other testing phases), but also whenever these repercussions could come into play in the future, e.g. when an operator in charge of a testing phase envisions that the device, process or framework being tested can carry ethical or legal repercussions once it is widely implemented.

For that to be the case, thorough training needs to be given to the operators carrying the task at hand, as well as to those with supervisory, organisational and managerial roles. For instance, an eye for sustainability and a 'culture of privacy' is often lacking in manufacturing plants that do not specialise in either, hence it is unlikely that any alarms would be raised in this regard. However, it is critical to incorporate these perspectives as early as possible in the implementation process to avoid the phenomenon of 'lock-in', both in terms of privacy and sustainability[1].

One must bear in mind that any health-related data is considered of sensitive nature under GDPR and thus the highest standards of safety apply. In the same vein, the AI Act considers any AI

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

applications that can have an effect on European citizens' employment to be of the highest category of risk. A device that records the operators' internal state, behaviour and conduct in real time can be used to evaluate the fitness and performance of individual employees to a detail and degree that was seldom possible in the past, hence it can gravely affect the access, promotion and review of the workforce. According to GDPR, and in order to instil trust in such data collection, it must be possible for operators to question the data, retrieve one's stored information, ask for a rectification and limit the range of data collected, as well as determine the purposes for which it is used. This is also crucial to ensure not only their free and willing cooperation, but also their safety, so as to avoid them being more preoccupied with the recorded parameters than focusing on their task and surroundings. However, the knowledge of the existence of such rights, as well as the practical ability to exercise them, is still anecdotal, due to the limited institutional and corporate efforts to make this informational available in a way that is adapted and comprehensible to the receiver.

Moreover, the scope of GDPR should be expanded so that employees are able to propose modifications to the process by which their data is stored, and to redirect the attention of the device to the parameters they deem most relevant to the task at hand. Further, given that the possibility of determining someone's degree of emotional arousal and task engagement, as well as the direction of visual attention, with such precision represents unchartered territory, there should be efforts to recognise the right to rest and the limitations of the human attention span and possibilities, to avoid placing overburdening expectations on operators.

The risk of violating the most intimate sphere of privacy calls for the further consideration of what has become known as 'neuro rights' (Unesco, 2022). Chile has become a pioneer in this field, by including in their new Constitution "legislation to protect mental privacy, free will and non-discrimination in citizens' access to neurotechnology", hence giving "personal brain data the same status as an organ, so that it cannot be bought or sold, trafficked or manipulated" (idem). The idea that some information would better stay outside of the scope of data collection such as emotional states and expressions (which come with a high degree of subjectivity, due to gender, culture, physiology, and many other contextual cues) should also be on the table, so that no aspect of data collection is left unquestioned, for which the importance of emphasising explainability and interpretability is key. Similarly, special attention must be paid whenever datasets are incomplete, unrepresentative of the wider makeup of the industry, or small.

Lastly, the inevitable limitations that come with the controlled environment of a Live Lab should not be understated. For instance, testing different operations in a chronological way instead of contemporarily is useful to analyse them individually, but sacrifices how realistic its effects are. Similarly, while in the context of a simulated environment, all values can be artificially maximised without friction. For instance, the experiments can abstract from any form of resistance coming from the operator, the wider industry or any other actor who might be resistant to change or widespread data collection.

[1] For more information on the concept of 'carbon lock-in', refer to section 3.7 of this Deliverable. In terms of data, the key issue to mention is that, given the data-hungry nature of real-time data-gathering devices, it is just as important to actively seek to minimise the collection of any data deemed unnecessary, risky or too invasive.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

## Live-Lab 2.1. Anomaly prediction and predictive maintenance. ESR 2. Devesh Jawla.

The sensor data collected from either machines or humans needs to vetted not only for ethical and legal reasons but also to ensure safety and accountability that all the critical data points are detected. The human in the loop which is required to teach the ML algorithms about the important bits of data will also work to ensure that critical data is handled with care. The human will ensure that the data is balanced or at least properly weighted so that the algorithm can work properly. This is due to the fact that almost every ML algorithm is sensitive to the class imbalance. This means that by nature every algorithm will be biased towards the majority of data samples.

The data we will collect at IVECO will be IoT and sensor data, these will be time series data which will need to be pre-processed. The pre-processing stage is unique for each problem and the data, and therefore we will have to carefully analyse the data and perform feature selection. These steps are typical for any machine learning project. For human performance review, the data needs to also be anonymised correctly and all ethical and legal guidelines need to be followed. This responsibility lies with each person who handles the data, from data collection to data pre-processing and ultimately data representation.

For IoT and Sensor Time-Series Data, anomaly detection uses AL, through uncertainty sampling. For example, suppose a highly imbalanced data with 999 regular events and only one abnormal event. In that case, the algorithm provides a label for this one abnormal event. There are usually only very few key events that require human attention in an exceptionally long time series. And therefore, to train a data driven anomaly detection model it may be necessary to review and label an exceptionally enormous amount of data points to identify enough anomalous datapoints for the data driven anomaly detection model to be trained on. This requires a human in the loop to teach and verify the ML algorithm for robustness, this kind of paradigm is known as Active learning. The idea is to iteratively train models on a task using tiny amounts of labelled data, use the model performance on the data to inform the selection of new data points for labelling, and then label the selected data points and retrain the model using the extended training dataset. The model is used to inform the selection of the data that are presented to a human annotator/expert for labelling and that are then used to retrain the model. AL finds applications in the industry due to a need for a safety critical system, and an efficient system so that it is practical for use in the industry. A Bayesian Deep Active learning framework for example provides credible predictions and therefore sits well with our requirements of a safety critical system. Moreover, in contrast to a passive ML system, we need to actively make sure that our system does not miss any valuable data. AL ensures this by way of uncertainty sampling and diversity sampling that we address all the classes, for example, if we have data which is highly imbalance with 999 normal events and only 1 abnormal event, we need an algorithm which makes sure that we provide a label for this 1 abnormal event. Secondly, AL naturally makes a supervised learning agent much more efficient because we require fewer samples to train our agent. IoT and Sensor Time series data can be explored much more efficiently by AL because usually in an exceptionally long time series there are only very few key events which require human attention. Here our task then becomes AL and Anomaly detection, where we first detect anomalies and then learn about only these major events, rather than learning about the complete time series.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

## Live-Lab 2.2. Legal and Ethical Implications. ESR10. Naira Lopez Cañellas

Given the further stage of development of this Live Lab, and the inclusion of legal and ethical provisions in its set up, a list of potential risks to look out for is provided below, so as to allow for a more dynamic, adjustable and straight-forward analysis of considerations to bear in mind. The list, although non-exhaustive, has been developed as a first draft, in order to continue to update it according to the needs of the Live Lab as it further progresses in its developmental phases. It has been primarily based on the Australian AI Work Health and Safety Scorecard (Cebulla et al., 2022).

The preliminary list of risks to kick-off an in-depth ethical and legal assessment of the Live Lab 2 reads as follows:

- Using AI when an alternative solution may be more appropriate/humane
- Displacing or undermining human decisions and capabilities
- Unintended consequences of false negatives and false positives
- AI being used out of scope or undermining social or company values
- Inequitable, exploitative or burdensome treatment of workers, or undermining their relations through gaming (also known as *reward hacking*)
- Attributing intelligence or empathy to AI system greater than appropriate
- Unfit/disproportionate collection of data
- Overconfidence in/overreliance on AI system, resulting in loss of/diminished due diligence or inability to react appropriately in the face of technical failure
- Security breach, data loss, cybersecurity vulnerability
- Impacting on other processes or essential services in the workplace
- Worker competencies and skills (not) meeting AI requirements
- Data collection (not) ceasing outside the workplace, retained longer than necessary or used for other purposes
- Insufficient update of assessment processes to review new AI-based operations
- Inadequate/insufficient specification or communication of purpose for AI use
- Little human oversight or checks and balances, so that algorithmic decisions and databases cannot be audited, challenged, contested, or improved
- Inadequate testing of AI in a production environment and/or for impact on different (target) populations.
- Helping widen inequalities and exacerbate biases against marginalised groups. Given that turning discrimination into measurable criteria is a challenging task, it is common to flatten variance within and among diverse sets of people (Tilmes, 2022). Similarly, it is hard to make sure to have representative sample/database/results, either when contrasting to a standard or seeking to create such standard (Cebulla et al., 2022)
- A workflow management system disproportionately, repeatedly or persistently assigning some workers to challenging tasks
- Failure to consider the environmental impact of one's activity and to opt for the most sustainable option (Brevini & Del Castillo, 2022)

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Once the risks have been laid out, it makes sense to propose some of the most relevant recommendations offered by the literature that can help answer them. For a more detailed discussion, refer to section 2.2. of this Deliverable. For a more schematic list, read below:

- Transparency on the location and ownership where data is stored, as well as regarding any third parties with access to it (Murdoch, 2021)
- Including all relevant stakeholders, proportionately to their degree of exposure to the device, in selecting the relevant metrics to record, and for which purposes (e.g. forbidding their use for the purposes of individual performance evaluation)
- Allocation of resources to developing the best suited methods of aggregation, anonymisation, encryption and protection of the data against cyber attacks, as well as for its periodic deletion (Todolí, 2019)
- Limitations to the monetisation of the data for purposes (or by actors) different than those initially stated (Murdoch, 2021)
- Actively practising data minimisation and decentralisation, so that it is stored locally and, whenever possible, data operations are conducted on a local level, as in the case of Federated Learning, Zero-Knowledge Proofs and Trusted Execution Environments (Gross et al., 2020)
- Helping develop Mediating Institutions for Data (MIDs) who advocate for the rights of the user base and are the ones to mediate and negotiate between them and the corporations that hold their data, as in the case of Data Unions (Gross et al., 2020)
- Any efforts to enhance transparency must include improving the receivers' digital literacy skills and ensure to give 'receiver-contextualised' explanations and statements of purpose, in order to avoid making the information too abstract, insufficient, tedious or unintelligible (Carpasso et al., 2022; Jiaobing & Gao, 2021)
- Any deployment of devices must happen incrementally and be embedded in an iterative process that allows for continuous feedback and improvement, and reports on any changes, technical or otherwise, which depart from the initial information given to stakeholders (Jiaobing & Gao, 2021)
- For the most pervasive AI applications, it is desirable to conduct an independent Human Rights Impact Assessments (HRIA) before their rollout, which may suggest a less intrusive course of action and even question the legitimacy of the purpose for its deployment (ECNL, 2021).
- *Wearables* are often subject to legislation that regulates commercial interests and the technological sector, rather than the stricter regulations applied traditionally to medical devices; therefore, it is crucial to anticipate their health-related implications and apply the scrutiny and pertinent safeguards that users expect for their health data (Segura Anaya et al., 2018; Gross et al., 2020)

## Live-Lab 2.3. Design experiments in an automotive sector line work environment for Human-machine performance monitoring and prediction. ESR 11. Carlos Albarran Morillo

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

For several reasons, it is essential to vet the sensor data collected from machines or humans carefully. First, there are ethical and legal considerations that need to be taken into account when collecting and using sensor data. For example, the data must be collected and used to respect individuals' privacy and comply with relevant laws and regulations.

In addition to ethical and legal considerations, it is essential to vet sensor data to ensure safety and accountability carefully. In safety-critical systems, all critical data points must be accurately detected and monitored to prevent accidents or errors. By carefully vetting the sensor data, we can ensure that all relevant data is collected and used and that the system operates safely and effectively.

Finally, carefully vetting sensor data can also help improve the data's accuracy and reliability. By carefully reviewing the data and ensuring that it meets specific quality standards, we can ensure that the data is reliable and can be used to make accurate predictions or decisions. Overall, carefully vetting sensor data is essential for ensuring sensor data's ethical, safe, and accountable use in various settings.

The data we will collect at IVECO will be IoT and sensor data; these will be time series data which is a type of data that records observations over time, often requires pre-processing before it can be used for machine learning tasks. This pre-processing stage is unique for each problem and data set and involves various steps designed to prepare the data for use in machine learning algorithms.

One common pre-processing step for time series data is feature selection, which involves identifying the most relevant and informative features for the task at hand. This can involve removing irrelevant or redundant features and transforming or combining existing ones to create new ones. By carefully selecting the most relevant and informative features, we can improve the performance of machine learning algorithms and reduce the risk of overfitting.

Another common pre-processing step for time series data is normalization, which involves scaling the data to have a consistent range and distribution. This can improve the performance of machine learning algorithms and ensure that the data is representative of the underlying phenomena being studied.

Overall, pre-processing is essential in any machine learning project and is particularly important for time series data. By carefully scrutinizing the data and completing the necessary pre-processing steps, we can improve the performance of machine learning algorithms and ensure that the data is ready for use in various tasks.

In addition, the data must be anonymized for human performance review. Anonymization is the process of removing personally identifiable information from data, making it impossible to identify individual people or organizations. This is often done for privacy and ethical reasons, as well as to comply with relevant laws and regulations. In the context of human performance review, anonymization is essential to ensure that individuals are not identifiable from the data being collected and analysed.

There are several ways to anonymize data, depending on the specific data set and the level of anonymity required. One common approach is to remove or replace sensitive information, such as names, addresses, and other identifying details. This can be done by replacing sensitive information with generic labels or codes or completely removing the information from the data set.

Another approach to anonymization is using techniques such as aggregation, perturbation, or suppression to make it difficult or impossible to identify individuals or organizations from the data.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

For example, data could be aggregated at a higher level, such as at the city or state level, to obscure the identities of individual people.

Overall, anonymization is an essential step in the process of human performance review, as it ensures that individuals are not identifiable from the data being collected and analyzed. By carefully anonymizing data, we can ensure that the data is used ethically and complies with relevant laws and regulations.

Wearable technology has made it possible to monitor an operator's internal state by tracking changes in their physiological signals. This can provide valuable information about an operator's level of fatigue, stress, or other factors affecting their performance and safety.

This live lab will collect data from the operators using a wristband (Fig 2.3.1) that collects physiological parameters and allows real-time data collection. The wristband can be equipped with sensors that collect data on the operator's physiological state and transmit this data wirelessly to a central computer or another device for analysis. This can provide real-time feedback to the operator, alerting them to changes in their physiological state and providing them with recommendations for improving their performance or reducing their risk of accidents. The data it provides:

The smartwatch comes with an advanced PPG sensor, an Accelerometer, an EDA sensor, and a Skin Temperature sensor to provide:

Raw sensor data (in AVRO Format, high-frequency raw data):

1. Accelerometer
2. EDA
3. Skin Temperature
4. Blood Volume Pulse (BVP)
5. Systolic peaks
6. On-wrist detection

Digital Biomarkers (in CSV format, the aggregated per minute digital biomarker CSV files contain the 1-minute physiological parameters calculated by Empatica’s proprietary algorithms)

- EDA
- Movement Intensity
- Pulse Rate (PR)
- Pulse Rate Variability (PRV)
- Respiratory Rate (RR)
- Skin Temperature

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901


Fig 2.3.1: EmbracePlus the wristband

Measurement of the physiological responses is less sensitive to participant bias and represents a viable alternative and accurate method to estimate human capabilities and task complexity. We can measure both factors using machine learning techniques. This device is not intrusive and can be used in a real-world application, not restricted to lab-based. We can monitor and predict the operator's behaviour using metabolic changes. Such wearable sensors can be worn while working, causing less operator discomfort. The vision of this live lab is to adopt an analytical approach to monitor workers' undesirable situations using physiological parameters. To analyse the patterns of physiological signals while workers perform different tasks, we will extract features employing statistical measures. The features that resulted in the highest unsafety situations prediction accuracy will be selected. Then, different classification algorithms will be utilized to construct a model that can predict unwanted states such as physical fatigue, stress, and mental workload in new cases. Finally, the presented method will help to recognize unsafety conditions more accurately. Overall, a comprehensive strategy for improving operator safety and performance in safety-critical systems can help to prevent accidents, develop warning systems against high levels of operator unwanted status, and design better operator-workstation allocation. This can help improve these systems' safety and efficiency and ultimately protect the lives and well-being of operators and other stakeholders.

## Live-Lab 3.1. Building a model for human performance prediction. ESR 1. Houda Briwa.

In many processes industry studies and reports, “human error” has been cited as a major contributing factor of incidents. The label” human error” is very controversial (Hollnagel, 1993) and the attribution of error is a judgment about human performance. Instead, it is more useful to investigate the cognition and behaviour of individuals and groups of people for reducing the potential for disaster in large, complex system (Woods et al., 1994). Labelling decisions and behaviours as errors highlight a symptom, not a root cause; the symptom should prompt a deeper examination into how a system

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

made up of individuals, groups of people, and technologies both works and fails ( (Hollnagel E. , Cognitive ergonomics and the reliability of cognition, 1991), (Hollnagel E. , Human reliability analysis., 1993), (Hollnagel E. , 2005), (Rasmussen, 1987), (Reason, 2000), (Sellen et al., 1989).

Typically, Human error probabilities (HEPs) are calculated using expert judgment, training simulator studies, or HRA data extrapolated from other sectors. However, the lack of adequate data has been highlighted as a major issue in the field of Human reliability assessment (HRA) (Park, 2019).

Further, despite many efforts in process safety assessment to provide a comprehensive model with the evidence-based causal relations between different Human factors issues and common circumstances affecting human performance or performance shaping factors (PSFs) in HRA, developing data-driven PSFs assessment models with enough transparency between models and source data, to deal with data scarcity and the uncertainty of expert judgments is still yet to be investigated for oil & gas operations (Liu, 2021). In this context, the Bayesian network (BN) has recently received the attention of researchers. In literature and in relation to HRA in control rooms, BN was applied for organizational factors‘ modelling, analysis of the relationships among failure influencing factors, assessment of human failure events, assessment of situation awareness and as an extension of existing HRA methods. An analysis of these research works has shown that the systematic investigation of all potential factor combinations or of the majority of combinations with appropriate simulator settings and the gathering of statistically meaningful data could be a very challenging objective to attend for many HRA models using PSFs. The analysis also demonstrated the need for more systematic frameworks to combine the various information sources pertinent to HRA (cognitive models, empirical data, and expert judgment) (Mkrtchyan, 2015).

Further, because of the vast differences between real operational conditions and simulator characteristics, validation of these models, as well as expert judgment models to generalize their applications and results, presents a set of challenges for HRA practitioners (Podofillini, 2020).

Simulators, through controlled experiments, provide a good opportunity to collect data for events that would not normally be seen in the historical record of process industries (such as severe accidents or unlikely PSF states like unavailable procedures) and allow to focus on specific causal factors to make inference on performance under known conditions. But the challenge of simulator data is accessibility: Most of simulator data are either not open to the public or have limited access to sensitivity of the raw data (Groth, 2014).

Furthermore, building a causal model that explicitly represents causal factors that impact the performance by combining cognitive studies, operating experience, simulator data can help alleviate the current challenges of the HRA field by providing both qualitative and quantitative information, regarding estimates of HEPs, correlations between various PSFs and their magnitude (Groth, 2014).

The first step to reach the goal of building such a model is to create a dataset that take into consideration existing HRA methods, expert judgment, and cognitive literature. In the Live Lab 3 experiment, we will be able to collect such data needed to implement the model. It will allow us to investigate the relationship between specific performance-shaping factors that contribute he human error and how they are impacting the operator in the event of alarm floods.

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

After investigation and an extensive literature review of the performance shaping factors used in different HRA models, a precise, and concise number of them will be selected as input of our model and different types of data will be collected in the experiment:

Sensor Data that will be collected using sensory devices such as EEG device and Eye tracker and while the participants are manipulating the simulator.

Subjective data that will be collect from the different Human factors assessment questionnaires. For instance, subjective methods to assess situation awareness of the participants during the experiment.

Expert knowledge data that will be the information gathered from the operating companies (partners in industry), either for parameterizing specific factors or to build the experiment based on real parameters. For example, Data regarding organizational factors, systems states, and interface characteristics

## Live-Lab 3.2. Predicting trips and critical scenarios using Bayesian Network (BN). ESR 3. Joseph Mietkiewicz.

My goal in Live Lab 3 as ESR 3 is to build a recommendation system for an operator in a control room to help in decision-making during a process disturbance. This recommendation system is based on an influence diagram. It is a probabilistic model that considers the probabilities of an event and its associated “cost”. The recommendation system recommends to the operator the action that reduces at most the "cost", in this case, the utility of the action. To do this, the recommendation system works in 3 steps: inference of the system state, prediction of the system state, and finally recommendation of the best action:

1. The system infers the state of the system based on the process data (temperature, pressure, etc.)
2. The system predicts the state of the system depending on the action of the operator
3. The system recommends to the operator the best action to take to have the best possible future state.

The error rate is considered in the recommendation to have a realistic prediction. A reinforcement learning agent is also trained. It takes into input the state of the system and recommends the best task to do. It takes a precise value to set to deal with the issue in the plant. The recommendation from reinforcement learning is also given to the operator.

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

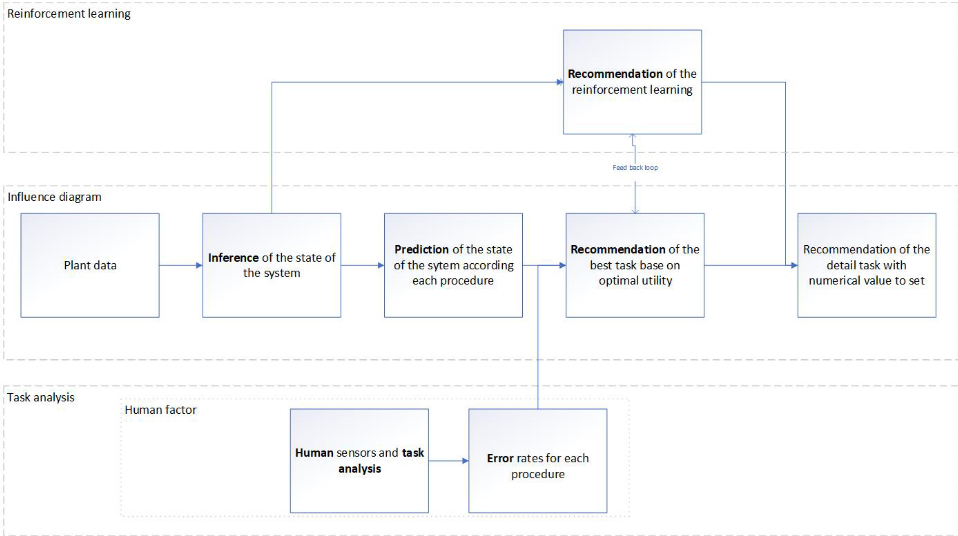



Figure 5.3.2.1: Graph of the functioning of the recommendation system

The model considers data coming from different sources. Process data coming directly from the simulator. It is data like pressure, temperature, … Those data are used to precisely predict the future state of the plant. This prediction involved expert knowledge in the form of physical equations and historical data. This part is the core of the model. This model includes other aspects and other types of data. Data links to task analysis are also considered. Those data include time available, probability of failure, … Those values come from a detailed analysis of each task. It is implemented in the model to give information to the operator about, for example how much time is left to do a specific task. Finally, data coming from sensors from the operator is considered in the model. The operator is equipped with sensors that provide information about his state regarding the work that he is doing. The sensor evaluates the workload of the operator. Thanks to this data error rates are calculated and implemented into the model. All in all, the model considers three types of data. Process data, task analysis data, and human data.

All in all, there are several key aspects of the recommendation system described in the provided information that are worth highlighting:

1. Data integration: By bringing data together from multiple sources into a single model, the operator can gain a more complete and accurate understanding of the state of the system. This includes data from the simulator (e.g., temperature, pressure), task analysis data (e.g., time available, probability of failure), and sensor data from the operator (e.g., workload, error rates).
2. Personalization: The recommendation system is designed to adapt to the operator and provide personalized recommendations based on the operator's current state and the task they are performing. This considers the unique characteristics of each operator and helps to ensure that the recommendations are relevant and appropriate for the operator's needs.
3. Accuracy and reliability: The recommendation system is designed to provide accurate and reliable recommendations that can help the operator to make informed decisions and improve the performance of the system. This includes considering expert knowledge in the

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

form of physical equations and historical data, as well as incorporating error rates based on sensor data from the operator.

4. Real-time decision-making: The recommendation system is intended to support the operator in real-time decision-making, providing recommendations and updates in real-time as new data becomes available. This can help the operator to respond quickly and effectively to changes in the system.
5. Prevention of accidents: By considering data from multiple sources and the human factor, the recommendation system can provide recommendations that can help to prevent accidents and improve the overall performance of the system.
6. Human-AI collaborative intelligence: The recommendation system is designed to be used in a human-AI collaborative intelligence context, with the aim of providing the operator with a comprehensive and personalized overview of the system. By considering the human factor and providing personalized recommendations, the system can better support the operator in their decision-making process.
7. Data security and privacy: Ensuring the security and privacy of the data being used in the recommendation system is critical. This includes protecting the data from unauthorized access and complying with the laws and regulations in place.
8. Data governance: Implementing effective data governance practices is important to ensure the integrity and reliability of the data being used in the recommendation system. This includes policies, procedures, and processes that ensure the responsible and effective use of data, as well as compliance with relevant data protection laws and regulations.

## Live-Lab 3.3. Process safety data modelling for human-in-the-loop configurations in process control. ESR 4. Chidera Winifred.

- Importance and need of data/data-integration with respect to performance monitoring/overview

Data is needed from the experimental study for the purpose of modelling the operators' behaviour during human-in-the-loop process control. The data collected would first be used to understand the interplay between latent factors and human, organisation, and technical factors in such configurations. Data collected would subsequently be used for human error probability estimation and operator performance monitoring. The identified critical factors can subsequently be addressed considering their impact to overall task risk. Also, Data collected from the simulator would be used to understand the systems performance.

- Importance and need of data/data integration with respect to preventing issues/accidents/failures

Live Lab 3.3 incorporates both human and system behaviour to address process safety concerns. Therefore, data will be collected from operators (in the experimental study) and the system log. The data would be used for probabilistically assessing and predicting task risk given the process safety case studies defined for Live Lab 3. This takes on an a priori approach to prevent process safety accidents.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

*o Importance and need of data/data-integration with respect to the integration of human in AI design*

Live Lab 3 covers the use of AI for risk modelling (incorporating the operators' behaviour). The model and methodological framework is to serve as a decision support tool for decision makers in safety critical status. The risk-based decision-making model and tool is to be developed using artificial intelligence techniques, algorithms, and tools. Tools such as Integrated dynamic decision analysis (IDDA), Bayesian network (BN) and potentially other machine learning algorithms will be applied to data collected from the experimental study.

*o Issues/challenges in data pre-processing and development (mostly in connection with the data integration in the context of human-AI collaborative intelligence)*

Challenges with respect to data integration can potentially be the integration of multi-modal data from the different data sources. Also merging and cross-referencing qualitative data with the quantitative data could potentially pose a challenge. However, these can be addressed with already existing strategies explored by other researchers.

*o Identify actions that need to be taken in connection to integrating the necessary data.*

The following actions are to be taken and are underway:

- Preparing for experiment (Design of experiment)
- Preparation of elements of interaction (simulated interface, procedures, monitoring tools, setup etc.)
- Recruitment of participants
- Experiment proper
- Database, data collection, analysis, and integration framework development
- Data collection using Eye tracker, heart rate sensors, surveys, sound recorder, and finally codes to record interfaces and human machine interaction activities, simulators/system log
- Single mode data pre-processing (cleaning, exploratory data analysis) of individual and group data, and cross verification of quantitative with qualitative data.
- Data transformation (if necessary)
- Code development for data integration
- Multi-modal data Integration and post-processing.

## Live-Lab 3.4. Process control optimization with alarm reduction and prioritization using Reinforcement Learning. ESR 5. Ammar Abbas.

*o Importance and need of data/data-integration with respect to performance monitoring/overview*

*Process optimization:* it is necessary to integrate the data in a processed format for an RL agent to be able to optimize the process.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

*Interpretations:* to provide interpretations to humans, the raw data does not provide meaningful information, however a processed data integration through Input-Output Hidden Markov Model (IOHMM) helps provide useful information in a human-understandable format.

o *Importance and need of data/data integration with respect to preventing issues/accidents/failures*

*Alarm Management:*

Reduction: raw data from sensors in the process industry is used by the IOHMM-DRL architecture from Figure 4.3.4.4 to reduce the number of alarms through process optimization and alarm management.

Prioritization: RL agent can help identify the most critical alarms from the sensor readings and reclassify them based on priorities by calculating the approximate future costs.

Sequencing: IOHMM can identify the sequence of alarms through correlation and clustering and help provide useful information for humans.

Suppression: RL agent can identify nuisance alarms based on the lower threshold for the criticality of the alarms and help suppress those alarms to help the operator for unneeded attention.

*Root Cause Analysis (RCA):* IOHMM can identify the root cause of the alarm and help humans to rectify the problem before the plant goes into emergency shutdown.

o *Importance and need of data/data-integration with respect to the integration of human in AI design*

*Human-AI collaborative agent:* Human-RL collaborative architecture is shown in Figure 5.3.4.1. The human is placed at the supervisory level and all the information is processed through the human agent. Therefore, it is necessary to first pre-process the raw sensor readings for the IOHMM and RL agent and then transform it to a human-understandable format.

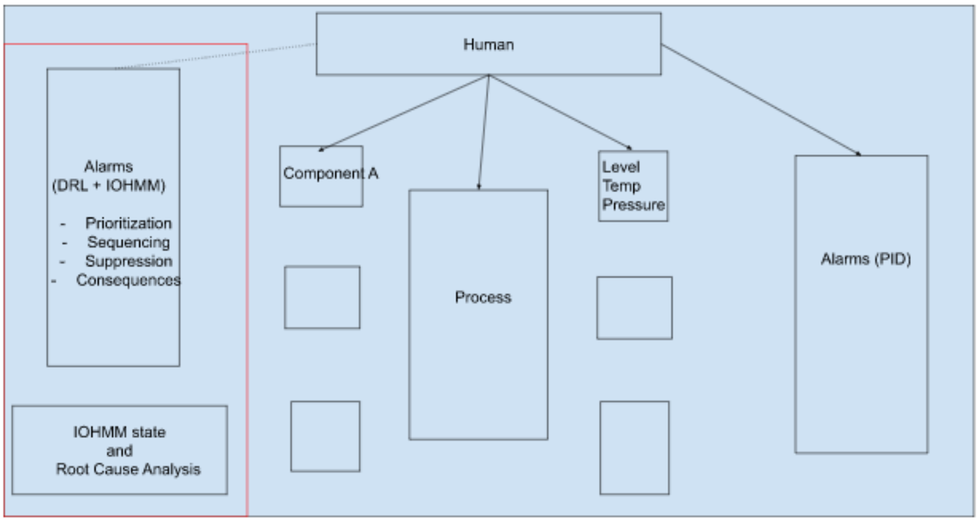



Figure 5.3.4.1: Flow diagram of Human-in-the-Loop RL design

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

*o Issues/challenges in data pre-processing and development (mostly in connection with the data integration in the context of human-AI collaborative intelligence)*

*Convergence Criteria:*

Convergence of IOHMM: It is essential for the IOHMM to converge to provide the necessary state information to the RL agent and for the interpretations provided to the human agent.

Convergence of Deep RL: convergence of the Deep RL is important to have an informed decision by the agent in safety-critical environments with near to zero chances of error.

*Assumptions:*

IOHMM is trained perfectly and provides the exact state information to the RL agent.

The data from the simulator/environment is synchronized without any missing data.

There is no lag between the system state information reading and the transmission of the control to the process.

*o Identify actions that need to be taken in connection to integrating the necessary data.*

*Data Pre-Processing:*

Principal Component Analysis (PCA): PCA is used for dimensionality reduction and provides the RL agent with the important sensor readings

MinMax normalization: min-max normalization is used for IOHMM for faster convergence and efficient learning

Standard normalization: Standard normalization is used for RL agent for faster convergence and efficient learning

Markov Modeling: IOHMM is based on the Markov model, and it is used as a form of feature engineering to provide interpretations and important state information to the RL agent.

*Requirements:*

Real-time interaction with the environment/simulator: a digital twin or interaction with the real system is necessary to train and test the Specialized Reinforcement Learning Agent (SRLA) (Abbas, Interpretable Input-Output Hidden Markov Model-Based Deep Reinforcement Learning for the Predictive Maintenance of Turbofan Engines, 2022).

Real-time human performance measurements: human performance measurement in real-time is important for the RL agent to be informed of the state of the human during stressful situations and for effective collaboration in a Human-AI environment

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Control inputs' and process variables' thresholds: information for the threshold is important for making an informed decision by SRLA.

Ability to introduce disturbance in the process: it is essential to be able to introduce synthetic known disturbance in the process for training the SRLA or IOHMM-DRL agent and testing the architecture.

Controller response during process disturbance: for pretraining the RL agent for faster convergence and minimum interaction with the environment it is important to have expert behaviour during the process disturbance.

## Live-Lab 3.5. Evaluation of the impact of simple/complicated visual assembly line instructions on operator's mental workload. ESR 6. Milos Pusica.

This experiment was designed to test the impact of complexity of task visual instructions on the worker's performance and workload. The experiment was conducted in laboratory conditions in modular and adaptive laboratory set-up for neuroergonomics and human-robot interaction research (Savković M, 2022). A worker was performing manual assembly task on small plexiglass plates, given multiple schemes images on a screen, and asked to assemble empty physical plates in the same way as shown on images (Figure 4.3.5.1). Schemes images were of two types - low complexity and high complexity schemes (Figure 4.3.5.2).

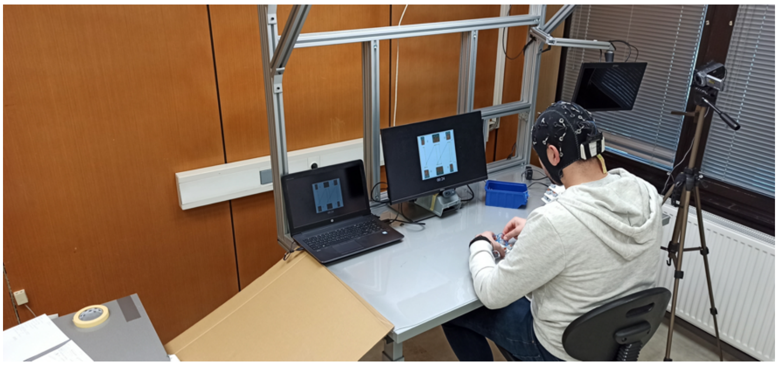


Figure 4.3.5.1 Worker performing manual assembly task on small plexiglass plates.

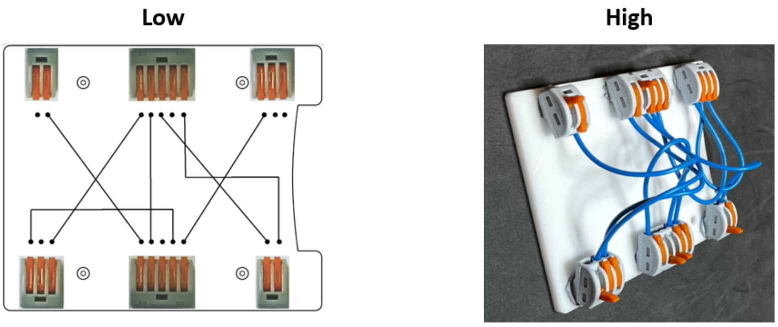



Figure 4.3.5.2 Schemes images of two types - low complexity and high complexity schemes

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

A worker is supposed to do the same task for both scheme types. However, high complexity schemes are designed to be more complex to understand and to cause more mental strain to the worker. Also, limited amount of time was given to each scheme, meaning that a new scheme image is given after the time expires or before, if the worker finishes the previous scheme earlier. Each experiment participant performed two sessions, each lasting for about 1.5h, with a short pause in-between.

For the whole duration of the experiment, EEG signals were recorded using Smarting Mobi, EEG-recording cap. Also, EMG signals of the participants were acquired. The experiment was recorded with two cameras, one side camera and one front camera, placed below the screen.

### Live-Lab 3.6. Using latent spaces to explore security challenging. ESR 9. Doaa Almhaithawi.

Security challenges solutions are vivid and under constant improvement due to the unprecedented and unpredicted intrusions, attacks, and even human errors. Although they usually have the same basic principles, they are dependent hugely on the data of the different scenarios. In our work, we are hoping to use the latent spaces methodologies and techniques across diverse situations, so far, there are multiple successful latent models' examples (ex: StyleGAN, Bert, Stable diffusion...) which had state-of-the-art results solving real-world problems in NLP, image processing, anomaly detection, ...

We started our work with building a new latent space using the simplest dataset possible (the natural numbers) to study the most evident features and priorities of this space while conducting multiple training attempts for the models with different setups, this will help us compare the different results and to investigate the effect of each setting on evident features (ex: odd and even). Next, depending on the obtained results, we will move to more complex spaces such as images, voice, video, and sensor readings, to solve security problems concerning anomaly, failure detection or pattern recognition.

### Live-Lab 3.7. Legal and Ethical Implications of Real-Time Data-Gathering Devices in the Workplace. ESR 10. Naira Lopez Cañellas.

Given that, at this stage, this is the least developed Live Lab, the overview provided below is also the most generalist. Its early stage of development offers a crucial opportunity to tackle its development process in the most holistic way, thus it promises to be a fertile ground for experimentation in the ethical and legal domains.

According to the literature reviewed to date the entrepreneurial drive to optimise the performance of operators in manufacturing sites has led to the development and deployment of a wide range of data-gathering devices in the workplace (Todolí-Signes, 2019). Tasked with creating, recording, processing, transferring and analysing data about a company’s operations, said devices store data about employee’s performance/characteristics as well as about the environment they interact with.

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

There is a vast potential benefit to deploying such instruments. These range from improving health, safety and security at work to optimising manufacturing activity and task allocation. Similarly, they can assist in finding the most efficient set up for a complex production line, as well as carrying out predictive maintenance and anomaly detection.

However, the presence of said technologies in the workplace can neither be neutral nor does it necessarily have a positive impact on all the stakeholders that interact with them (Murdoch, 2021). Depending on the choice of device, and the nature of their implementation process, they can pose a risk to a myriad of consolidated workplace and fundamental rights. Consequently, the design, development and assessment of the Live Lab needs to include a legal and ethical perspective so that socioeconomic, political, cultural and environmental factors are taken into account.

There are two primary ways in which these factors enter the discussion. On the one hand, legal frameworks play a central role. On a European level, the most salient ones include GDPR, the Data Act, the Data Governance Act, the AI Act, and European charters and treaties regarding fundamental and socioeconomic rights. On a national level, the relevant legislation becomes even more specific, regulating matters such as sustainability and labour rights, in addition to all the sector-specific health and safety regulations. On the other hand, ethics comes into the picture to complement legislation. The list of relevant documents in this regard is also considerable: The Montreal Declaration for a Responsible Development of Artificial Intelligence (2017), UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021), the extensive work done by the Council of Europe (CoE) Ad hoc Committee on Artificial Intelligence (CAHAI, 2021), the Ethics Guidelines for Trustworthy AI by the EU High-Level Expert Group on Artificial Intelligence (HLEG AI, 2020), or the Data Ethics Decision Aid (DEDA) by the Utrecht Data School (2021)[1].

However, these can often be too abstract, or removed from their contexts of application. In order to translate them into the case of this Live Lab (and, similarly, to the case of any other manufacturing facility with a comparable degree of human-robot collaboration) more granular tools are needed. Such tools include the Australian AI Work Health and Safety Scorecard (Cebulla et al., 2022: 2), which helps 'assess and manage the potential risks and hazards to workers resulting from AI use in the workplace'. These are in turn identified by pooling together the Australian AI Ethics Principles and the Principles of Good Work Design, and contextualising them in the different stages of AI implementation as devised in the AI Canvas (ibidem). The main elements of the analysis are as follows:

- Different stages of AI implementation, as found in the AI Canvas (Cebulla et al., 2022: 6): prediction, judgement, action, outcome, training, input, feedback
- Australian AI Ethics Principles (Cebulla et al., 2022: 7): human, social and environmental wellbeing; human-centred values; fairness; privacy and security; reliability and safety; transparency and explainability; contestability; accountability
- Key Characteristics of Work, as found in the Principles of Good Work Design (Cebulla et al., 2022: 4): Physical (including chemical, physical and biological hazards); Cognitive (e.g. information processing load, complexity and duration); Biomechanical (which mainly refers, in turn, to force, movement, posture, vibration) and Psychological (which stretches from the need for the support by peers and supervisors, to managing one's expectations at work, or being free from stress, overwork and fatigue)

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Therefore, the scorecard provides a springboard from which to start incorporating a regulatory and ethical point of view to the design of this LiveLab. The assessment starts off with the framework provided by this scorecard, and as the analysis progresses and the insights of multiple stakeholders are combined, it will evolve by highlighting its limitations, and whenever necessary, propose alternatives better suited to the context of the Live Lab. While the first and most straight-forward benefit of applying this tool is that it helps link a wide range of workplace hazards to ethics principles and to the different stages of the AI lifecycle in a company, there are a few more advantages worth pointing out.

A second benefit of the application of this scorecard is to help reconcile the inevitable normative trade-offs when deploying such devices, as all objectives cannot be maximised at once. That is the case, for instance, with regard to technical accuracy vis-à-vis human interpretability (Izumo & Weng, 2022). Similarly, higher efficiency can come at the expense of enjoyment of the workplace activity, more natural workflows, or workers' self-organisation and decision-making (Cebulla, 2022). In this context, competing values are at play, such as when higher levels of autonomy, transparency and democratisation entails higher requirements in terms of skillset, which can be resource-intensive and/or enlarge inequalities between different members of the staff. Finally, it will also hint at the tensions that are bound to arise when different stakeholder interests may be at odds with each other, e.g. when shareholder's preference for higher productivity requires shrinking the workforce or modifying their usual tasks (Izumo & Weng, 2022).

Furthermore, the scorecard allows us to focus on the severe socio-ecological impacts of AI, often overlooked in manufacturing contexts like the ones the Live Labs seek to emulate. Its reference to the principle of "human, social and environmental wellbeing" seeks to account for AI's role in intense resource and energy consumption, creation of e-waste, pollution of natural resources and extraction of toxic materials (Brevini & del Castillo, 2022; Murdoch, 2017). The case of lithium is one of the most renowned, as its use in the EU is projected to rise meteorically (by 3500% according to the EC's 2022 estimates), but its extraction comes at a huge cost. In Chile, the world's 2nd largest producer, mining has severely damaged water access, indigenous lands and biodiversity (Brevini & Del Castillo, 2022). While lithium-ion batteries can be recycled, it remains a marginal practice, due to a mix of 'technical constraints, economic barriers, logistic issues, and regulatory gaps' (Jacoby, 2019), and an overall lack of interest and investment in prioritising re-use over new mineral exploitation.

Similarly, the data centres that many AI applications rely on for training and use require large amounts of energy and water that are equivalent to the use of entire countries as large as Iran (Hao, 2019). This is compounded by the rebound effect at play wherever AI is applied in the oil and gas industries, as the gains in the efficiency of their processing often lead to an increase in their overall extraction. To make matters worse, most AI applications require particularly high levels of electricity consumption, which given that two thirds of the world's electricity grid is still based on fossil fuels, presents quite an environmental challenge in and of itself. Finally, it is equally unsettling to recognise it is still common practice to store material on the cloud without knowledge of, or regard to, its environmental impact (Brevini & Del Castillo, 2022). It is worth pointing out that the most harmful of these processes primarily take place in the Global South, as do the most aggressive forms of labour and data extractivism (idem).

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Hence, the scorecard allows to negatively grade those devices that do not seek to prevent, mitigate or compensate for their social and environmental footprint. It can also be used to develop labels and traceability methods that improve transparency in this regard (Cebulla, 2022). The urgency of adopting such measures has to do with the fact that AI is becoming an infrastructure supporting increasing levels of service provision (Robbins & van Wynsberghe, 2022). It is therefore of prime importance that these tools are developed prior to widespread AI adoption, so that they can be embedded in its implementation process as it rolls out. A failure to do so would result in carbon lock-in, at which point the dependency on carbon is hard to reverse for technologic, economic, political and social reasons.

However, and most centrally, the main strength of the scorecard approach is that it puts the complicated relationship between *wearables* and labour rights at the centre of the discussion. While the use of *wearables* holds great potential to minimise some of the most renowned risks in a manufacturing site (e.g. avoiding collisions, information overload, physical injuries and fatigue) it can also be counterproductive (due to newly created distractions or disproportionately high performance expectations) or even generate new ones (especially in the realm of privacy, cybersecurity, data and personal fulfilment at work) (Todolí-Signes, 2020).

The data-related risks include a high degree of intrusiveness in the data that is collected, often beyond what is strictly necessary and stretching outside the boundaries of the workplace. The quality of the protection against misuse or access by third parties is often poor, as is workers' awareness and training on their rights regarding refusal, access, modification, portability or erasure. There are hardly any mechanisms to prevent using such data for discriminatory purposes – including recruiting, hiring, promoting and firing employees –, or to ensure the exposure to said data-collecting mechanisms is equal throughout a manufacturing plant (Todolí-Signes, 2019).

What's more, acquiring consent for the use of data-gathering devices in the workplace is further complicated by the relation of subordination between employer and employee, which removes the option to opt out freely or to enforce one's rights (Ponce, 2021). Along the same line, existing avenues for redress in the face of any arbitrary or significantly harmful impact on individual employees are far from homogenous, and rely mostly on the discretion of each individual private actor or institution (Wildhaber, 2018).

The list of risks related to one's cognitive and psychological needs is also extensive, and includes an overreliance on technology, a significant loss of skills or attention span, a disproportionate surge in expected productivity or risk-taking behaviour, or a loss of workers' autonomy, which hampers dignity and meaning at work (Cebulla et al., 2022). Todolí-Signes (2019) also sheds light on the adverse effects on workers' mental health of receiving 'mental whips' when they take pauses or act erroneously; of recording data that risks infringing their fundamental rights (such as their location or interactions throughout the day), or emotion-recognition and biometric identification technology, even if only devised with the purpose of recognising when a worker is too tired or in need of assistance.

Finally, it is crucial to interrogate and problematise the values that often underpin the most renowned charters and codes of ethics, as their selection, definition, and underlying tensions are often overlooked. The drafting of guidelines containing ethical principles and values, although crucial as a backbone to inform and ensure coherence within and between codes of conduct and legislation, are

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

only the first step of the process. Similarly, an agreement on the language and the hierarchy of said values is insufficient to ensure their fulfilment can be monitored in practice. On the one hand, there is a considerable degree of institutional agreement on a key set of values on a European level, around varying combinations including trustworthiness, accountability, transparency, safety, fairness, confidentiality, etc. (CAHAI 2021; HLEG AI, 2020). Far from shying away from these complications, Live Labs offer the perfect environment to test them in a safe environment, and to try out different arrangements and definitions that garner the highest degree of approval among those who will be most affected by the workplace reconfiguration that *wearables* will entail.

An interesting piece of work that can serve as inspiration in this regard is an experiment by Schoeffer et al. (2022) which concludes with the need to *decenter* explainability as the main tool to test the likelihood of human override of AI decisions. Their findings show that the most relevant predictor for human intervention in AI rulings, over the degree of explainability or even truthfulness of the outcome, is the perception of fairness of the AI-originated recommendation.

An even more telling case can be found in the work of (Andrada et al., 2022), who go one step further in their reasoning and show that different conceptualisations of the same value (in this case, transparency) can be at odds with each other. On the one hand, *reflective transparency* refers to users' ability to break down and understand the technical workings of the AI application. This helps them make sense of the lifecycle of the materials it is made of, and the impact it has on human behaviour and cognition. On the other hand, *transparency-in-use* refers to a generalised preference for such intuitive designs that users do not need to reflect on the device and can focus instead on the tasks it is made to help them with. Hence, the two concepts are in tension with each other, as a higher degree of knowledge and mastery of a device usually makes the user (or developer) gradually less and less attuned to any 'bias to our cognitive processes in hidden ways' or the ways in which they 'jeopardize our selfhood or intellectual autonomy (idem: 6). And yet, technological use and innovation unequivocally seeks to make devices less obscure and burdensome to use, as this is one of the main drivers of their widespread adoption.

Overall, this section has outlined the specific ethical and legal issues raised by the deployment of wearables and AI driven data gathering and analysis in a manufacturing site.. As Live Lab 3 advances, more concrete research experiments and proposals will develop. For a schematic delivery of some of the main recommendations offered in the literature in order to tackle the risks that have been highlighted, refer to section 2.2. of this Deliverable.

[1] The following two repositories offer a comprehensive selection of some of the core documents in the field: Council of Europe (n.d.). *Ethical frameworks*. Available online at: https://www.coe.int/en/web/artificial-intelligence/ethical-frameworks

GAIEC (n.d.). *GAIEC Repository*, Global AI Ethics Consortium. Available at: https://www.ieai.sot.tum.de/global-ai-ethics-consortium/gaiec-repository/

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

## Live-Lab 3.8. Testing of different ML approaches for EEG feature extraction to detect operator's mental workload. ESR14. Andres Alonso Perez.

The goal of the ESR14 would be to develop an Artificial Intelligence model capable of extracting relevant EEG features and test it to assess Mental Workload and for multiple applications inside Human Factors.

- *Importance and need of data/data-integration with respect to performance monitoring/overview:*

When collecting the Mental Workload and the Situational Awareness, the performance of the operators can be related to them (e.g., the correlation between low performance and high levels of workload can be investigated) and, with the task's allocation, the performance could, potentially, increase. The Situational Awareness gives a measure of the understanding and the knowledge an operator has of the situation and could be used as a measure of the performance.

Nevertheless, this part of the project is more oriented to assess the state of the worker rather than to measure the performance. Other parts of Live Lab 3 like live lab 3.1 directly address this.

- *Importance and need of data/data-integration with respect to preventing issues/accidents/failures:*

This part of the project, as stated, has the purpose of being able to assess in real time the state of the worker. This, along with the task's allocation, has the potential to put the operator in the best situation possible, preventing critical states of the worker like an excessive workload or a low Situational Awareness by correctly designing the working environment, or raising an alarm when the operator is being saturated. This can be a way to prevent accidents in the working space from its very conception.

- *Issues/challenges in data pre-processing and development (mostly in connection with the data integration in the context of human-AI collaborative intelligence)*

The main issues here would be the collection of the data since EEG requires very specific and noise-free environments and can be uncomfortable for the developing of certain tasks. Also, one of the biggest challenges is that the model that assesses mental workload should be able to work in real time. So, not only there should be a correct streaming of data, but the model Artificial Intelligence that assigns a workload should be compact enough to have a low inference time so there is not a significant delay between the data input and the mental workload estimation. Therefore, there would be a trade-off between the mental workload and the accuracy of the model.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

# 4. Conclusions

In this deliverable the types of data sets as well as the requirements for data collections are described for each LiveLab. As explained before, the LiveLabs have different focus, however, they are all connected to each other. Therefore, the value of the consortium resides in the high level of collaboration between the ESRs, their supervisors and the network in general. Some ESRs have already started the design of their experiments for data collection. The following preparations have been made.

LiveLab 1 will focus on the implications and methodologies involved in the human-robot collaboration. This includes the analysis of the physiological state of the operator during the collaborative tasks through advanced neuroergonomics experiments, the study the factors that affect the performance of operator teleoperating of robots, an interactive programming method for robots which transfers human skills to robots through demonstration, and the safety of the operator in task demonstration as well as the assessment of human performance in kinaesthetic teaching.

In LiveLab 2 ESRs are involved in the operator performance in a manufacturing operation environment. For this LiveLab, an integral approach is taken as for the need of sensory data from the machinery interaction (such as welding operations), the performance (parts completed without errors, unsafe conditions etc.), and environmental conditions. The risks and recommendations involved in the data collection processes are being formulated according to the literature available. It has been pointed out that it is essential to vet the sensor data collected from machines or humans carefully.

Live Lab 3 focuses on researching and addressing challenges in alarm handling and process control design and management in safety-critical systems, with a focus on human and system behaviour during alarm management and human-in-the-loop process control. The research team, consisting of ESRs, will use a simulated formaldehyde production facility as a case study and conduct simulations of three increasingly complex scenarios involving failures in the pressure control and cooling systems. The ESRs will collect data on operator and system behaviour during these simulations and analyse the impact of developed support systems on operator behaviour and performance. The overall goal of the research is to develop a framework and tools for decision support in process safety critical situations and for optimizing process control and process safety in human-in-the-loop configurations. This will involve modelling human performance and decision-making in critical scenarios, predicting trips and critical scenarios using Bayesian networks, and modelling process safety data in human-in-the-loop configurations.

Overall, there are ethical and legal considerations that need to be considered when collecting and using sensor or human-related data. For example, the data must be collected and used to respect individuals' privacy and comply with relevant laws and regulations. In addition to ethical and legal considerations, it is essential to vet sensor data to ensure safety and accountability carefully. In safety-critical systems, all critical data points must be accurately detected and monitored to prevent accidents or errors. By carefully vetting the sensor data, we can ensure that all relevant data is collected and used, and that the system operates safely and effectively.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

## 4.1. Future actions

According to the descriptions provided in Section 3, the proposed modular workstation in the LiveLab 1 gives the go-ahead to perform neuroergonomic experiments in manufacturing scenarios, involving state-of-the-art systems and sensors supporting the operator to accomplish the tasks. During the experiments, muscle and brain activity are monitored in real time during the performance of repetitive, monotonous assembly activities by EEG and EMG sensors. The collection of these data is the input to analyse the physiological performance of the operator during the tasks. For the Live Lab 1.2 a mobile EEG cap from mBrainTrain (Smarting mobi, mBrainTrain, Serbia) be used with a sampling rate range between 250Hz-500Hz and a Tobii Pro Nano eye-tracker mounted on a monitor, with sampling rate of 60Hz.

For the teaching by demonstration process for robots an interactive programming method for robots, an ITL framework has been proposed which transfers human skills to robots through demonstration. Then, a deep neural-network-based anomaly/failure detection module is deployed as a safeguard to prompt a user in safety-critical events. Subsequently, we incrementally refine the system's skill library by online learning from successful and failed executions. This experiments will be carried out in a collaborative robot cell (Intelligent Motion and Robotics) technology. The cell consists of a UR10 robot arm with a gravity compensation controller, allowing it to perform kinaesthetic tasks through human demonstrations. This experiment will also help to assess the safety of the operator in task demonstration as well as the assessment of human performance in kinaesthetic teaching. The project will start with the risk assessment the risk assessment paradigm which addressed the use of hybrid Standarization format and focuses on the functional safety of cobot cell. The objective is to make safe and reliable human robot interaction and the focal standard in our application will be ISO 10218 and ISO/TS 15066 standards with other normative standards. The risk assessment will be performed using Pilz safety assessment protocols.

For the LiveLab 2 at IVECO the data collection will be IoT and sensor data, these will be time series data which will need to be pre-processed. In addition to that, the LiveLab 2 will gather data from the operators using a wristband that collects physiological parameters and allows real-time data collection. The wristband can be equipped with sensors that collect data on the operator's physiological state and transmit this data wirelessly to a central computer or another device for analysis. However, talks with the IVECO part are still in place to determine the best way to carry out the data collection.

Based on the provided information, some potential future actions for Live Lab 3 might include: Conducting experimental studies using the simulated formaldehyde production facility to study human and system behaviour during alarm management and human-in-the-loop process control. Testing developed support systems for their impact on operator behaviour and performance during the simulations. Building a model for human performance prediction in critical scenarios using a Bayesian network. Designing an experiment for HRA (Human Reliability assessment) data collection. Predicting trips and critical scenarios using Bayesian networks to assist decision-making for a human operator and comparing different approaches to building the networks. Modelling process safety data in human-in-the-loop configurations. Developing an innovative methodological framework and tools for decision support in process safety critical situations and for optimizing process control and process

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

safety in human-in-the-loop configurations. Conducting additional research and analysis to address the concerns of Live Lab 3.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

# References

Abbas, A. N. (2022). Deep Residual Policy Reinforcement Learning as a Corrective Term in Process Control for Alarm Reduction: A Preliminary Report. *European Conference on Safety and Reliability (ESREL)* (pp. 3260-3266). Dublin, Ireland: RESEARCH PUBLISHING, SINGAPORE.

Abbas, A. N. (2022). Interpretable Input-Output Hidden Markov Model-Based Deep Reinforcement Learning for the Predictive Maintenance of Turbofan Engines. *In International Conference on Big Data Analytics and Knowledge Discovery (DaWaK).* Springer Cham.

Ahn, Sangtae, Thien Nguyen, Hyojung Jang, Jae G. Kim, and Sung C. Jun. 2016. "Exploring Neuro-Physiological Correlates of Drivers' Mental Fatigue Caused by Sleep Deprivation Using Simultaneous EEG, ECG, and FNIRS Data." Frontiers in Human Neuroscience 0 (MAY2016): 219. https://doi.org/10.3389/FNHUM.2016.00219.

Albarrán Morillo, C.; Jawla, D.; D. Kelleher, J.; Demichela, M. (2022). "Safety-critical systems in the automotive sector: pros and cons in the current state-of-the-art of human performance assessment". Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)

Alwassel, Humam, Dhruv Mahajan, Bruno Korbar, Lorenzo Torresani, Bernard Ghanem, and Du Tran. 2019. "Self-Supervised Learning by Cross-Modal Audio-Video Clustering." Advances in Neural Information Processing Systems 2020-December (November). https://doi.org/10.48550/arxiv.1911.12667.

Andrada, G., Clowes, R.W. & Smart, P.R (2022). Varieties of transparency: exploring agency within AI systems. AI & Soc. Available at: https://doi.org/10.1007/s00146-021-01326-6.

Antonenko, P.D., Paas, F., Grabner, R.H., & Gog, T.V. (2010). Using Electroencephalography to Measure Cognitive Load. Educational Psychology Review, 22, 425-438.

B. Taskar. Learning structured prediction models: a large margin approach. PhD thesis, Stanford University, 2005.

Bakshi, V.K. (2018). Towards Practical Driver Cognitive Workload Monitoring via Electroencephalography.

Berger, Suzanne, and Benjamin Armstrong. "The Puzzle of the Missing Robots." (2022).

Bosch, T., de Looze, M. P., van Dieën, J. H., (2007). Development of fatigue and discomfort inthe upper trapezius muscle during light manual work. Ergonomics, 50(2), 161-177.

Bradley, Margaret M., Laura Miccoli, Miguel A. Escrig, and Peter J. Lang. 2008. "The Pupil as a Measure of Emotional Arousal and Autonomic Activation." Psychophysiology 45 (4): 602–7. https://doi.org/10.1111/J.1469-8986.2008.00654.X.

Brevini, B. & Ponce Del Castillo, B. (2022, July 5). Artificial intelligence, artificial solutions: Why the environmental impact of AI is overlooked in contemporary public debates?, by the European

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Trade Union Institute (ETUI), available online at: https://www.etui.org/events/ai-talks-etui-artificial-intelligence-artificial-solutions

Brookings, J.B., Wilson, G.F., & Swain, C.R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. Biological Psychology, 42, 361-377.

Brouwer, A., Zander, T.O., van Erp, J.B., Korteling, J.E., & Bronkhorst, A.W. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. Frontiers in Neuroscience, 9.

Capasso, M. & Umbrello, S. (2022). Responsible nudging for social good: new healthcare skills for AI-driven digital personal assistants. Med Health Care and Philos 25, 11–22. Available at: https://doi.org/10.1007/s11019-021-10062-zCeh, Simon M., Sonja Annerer-Walcher, Christof Körner, Christian Rominger, Silvia E. Kober, Andreas Fink, and Mathias Benedek. 2019. "Data for: Neurophysiological Indicators of Internal Attention: An EEG-Eye-Tracking Co-Registration Study." OSF. 2019. https://doi.org/10.17605/OSF.IO/5U6R9.

Cebulla, A. et al. (2022). Applying ethics to AI in the workplace: the design of a scorecard for Australian workplace health and safety. AI & society, 1–17. Available at: https://doi.org/10.1007/s00146-022-01460-9

Charles, R., Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. Applied ergonomics, 74, 221-232.

Chernova, Sonia, and Andrea L. Thomaz. "Robot learning from human teachers." Synthesis lectures on artificial intelligence and machine learning 8, no. 3 (2014): 1-121.

Colledanchise, Michele, and Lorenzo Natale. "On the implementation of behavior trees in robotics." IEEE Robotics and Automation Letters 6, no. 3 (2021): 5929-5936.

Comberti, L., Demichela, M., & Leva, M. C. A multi-discipline method to assess the human performance in manufacturing industry for safety and quality optimization. In Safety and Reliability–Safe Societies in a Changing World (pp. 381-386). CRC Press. (2018).

Coster, F. (2017). Smart workload balancing. *H-Workload 2017: The first international symposium on human mental workload Dublin Institute of Technology.*

Cuninka, P., & Strémy, M. (2015). Evaluation of Safety Parameters According to IEC Standards. *Research Papers Faculty of Materials Science and Technology Slovak University of Technology*, *23*(s1), 29–36. https://doi.org/10.1515/rput-2015-0024

DATA PROTECTION COMMISSION. (2021). *Definition Key Terms*. Retrieved from https://www.dataprotection.ie/en/individuals/data-protection-basics/definition-key-terms

Dehais, Frédéric, Alban Duprès, Gianluca Di Flumeri, Kevin Verdière, Gianluca Borghini, Fabio Babiloni, and Raphaëlle Roy. 2019. "Monitoring Pilot's Cognitive Fatigue with Engagement Features in Simulated and Actual Flight Conditions Using an Hybrid FNIRS-EEG Passive BCI." Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018, January, 544–49. https://doi.org/10.1109/SMC.2018.00102.

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Dehais, F., Lafont, A., Roy, R., & Fairclough, S. (2020). A neuroergonomics approach to mental workload, engagement and human performance. *Frontiers in neuroscience, 14*, 268.

Dessai, Amita, and Hassanali Virani. 2022. "Emotion Detection Using Physiological Signals." In International Conference on Electrical, Computer, and Energy Technologies, ICECET 2021, 1–4. https://doi.org/10.1109/icecet52533.2021.9698729.

Doshi, K., & Yilmaz, Y. (2022). Zero-Shot Action Recognition with Transformer-based Video Semantic Embedding. http://arxiv.org/abs/2203.05156

European Commission (2000). Communication from the Commission on the precautionary principle. Available online at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52000DC0001

European Commission (2019). What is GDPR, the EU's new data protection law? GDPR.EU. Available online at: https://gdpr.eu/what-is-gdpr/

European Comission: Directorate-General for Migration and Home Affairs. (2020). *Guidance note — Potential misuse of research.* Retrieved December 2021, from https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/guide_research-misuse_en.pdf

European Union Commission. (2021, June 4). *on standard contractual clauses for the transfer of personal data to third countries pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council*. (COMMISSION IMPLEMENTING DECISION (EU) 2021/914) Retrieved December 2021, from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32021D0914&from=FR

ECNL (2021, March 23). Evaluating the risk of AI systems to human rights from a tier-based approach. European Centre for Not-For-Profit Law. Available at: https://ecnl.org/news/evaluating-risk-ai-systems-human-rights-tier-based-approach

Gross, M.S. et al. (2020). 18th White Paper on Ethical Implementation of Wearables in Pandemic Response: A Call for a Paradigm Shift. Ethics Harvard, Edmond J. Safra Center for Ethics. https://ethics.harvard.edu/sites/hwpi.harvard.edu/files/center-for-ethics/files/18ethicalwearables.pdf?m=1590163395Gevins, A.S., Leong, H.M., Du, R., Smith, M.E., Le, J., DuRousseau, D.R., Zhang, J., & Libove, J. (1995). Towards measurement of brain function in operational environments. Biological Psychology, 40, 169-186.

Groth, K. M. (2014). A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods. *Reliability Engineering & System Safety, 128*, 32-40.

Guzmán, L. H. (2022). Chile: Pioneering the protection of neurorights. The UNESCO Courier. Available online at: https://en.unesco.org/courier/2022-1/chile-pioneering-protection-neurorights

H. Daumé III, J. Langford, and D. Marcu. Search-based structured prediction. Machine Learning, 75:297–325, 2009.


This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Hao, K. (2019, June 6). Training a single AI model can emit as much carbon as five cars in their lifetimes. MIT Technology Review. https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/He, J., Zhu, X., Combining Improved Gray-Level Co-Occurrence Matrix With High Density Grid for Myoelectric Control Robustness to Electrode Shift (2017).IEEE Trans. Neural Syst. Rehabil., 25, 1539-1548.

Healey, Jennifer A., and Rosalind W. Picard. 2005. "Detecting Stress during Real-World Driving Tasks Using Physiological Sensors." IEEE Transactions on Intelligent Transportation Systems 6 (2): 156–66. https://doi.org/10.1109/TITS.2005.848368.

Hiatt, L. M., Narber, C., Bekele, E., Khemlani, S. S., & Trafton, J. G. (2017). Human modeling for human–robot collaboration. *International Journal of Robotics Research*, *36*(5–7), 580–596. https://doi.org/10.1177/0278364917690592

HLEGAI (2019). Ethics guidelines for trustworthy AI. High Level Expert Group on Artificial Intelligence. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Hollnagel, E. (1991). Cognitive ergonomics and the reliability of cognition. *Le Travail Humain* , 305-321.

Hollnagel, E. (1993). The phenotype of erroneous actions. *International Journal of Man-Machine Studies*, *39*(1). https://doi.org/10.1006/imms.1993.1051

Hollnagel, E. (2005). Human reliability assessment in context. . *Nuclear engineering and technology 37.2* , 159-166.

Holm, A., Lukander, K., Korpela, J., Sallinen, M., & Müller, K. M. (2009). Estimating brain load from the EEG. *The Scientific World Journal, 9*, 639-651.

Ijspeert, Auke Jan, Jun Nakanishi, Heiko Hoffmann, Peter Pastor, and Stefan Schaal. "Dynamical movement primitives: learning attractor models for motor behaviors." Neural computation 25, no. 2 (2013): 328-373.

Izumo, T., & Weng, Y. H. (2022). Coarse ethics: How to ethically assess explainable artificial intelligence. AI and Ethics, 2(3), 449–461. https://doi.org/10.1007/s43681-021-00091-y

Jacoby, M. (2019). It's time to get serious about recycling lithium-ion batteries. Chemical and Engineering News. Available online at: https://cen.acs.org/materials/energy-storage/time-serious-recycling-lithium/97/i28

Jiménez, Sergio, Tomás De La Rosa, Susana Fernández, Fernando Fernández, and Daniel Borrajo. "A review of machine learning for automated planning." The Knowledge Engineering Review 27, no. 4 (2012): 433-467.

Khan, I. U., Afzal, S., & Lee, J. W. (2022). Human activity recognition via hybrid deep learning based model. *Sensors*, *22*(1). https://doi.org/10.3390/s22010323

This project has received funding from the European Union's Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Laird, John E., Kevin Gluck, John Anderson, Kenneth D. Forbus, Odest Chadwicke Jenkins, Christian Lebiere, Dario Salvucci et al. "Interactive task learning." IEEE Intelligent Systems 32, no. 4 (2017): 6-21.

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering, 15*(5), 056013.

Lentini, Gianluca, Giorgio Grioli, Manuel G. Catalano, and Antonio Bicchi. "Robot programming without coding." In 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 7576-7582. IEEE, 2020.

Leva, M. C., Alunni, C. C., Demichela, M., & Allemandi, G. “Addressing human performance in automotive industry: identifying main drivers of human reliability”. Irish Ergonomics Society, 18 (2016).

Leva, M. C., Demichela, M., Comberti, L., & Caimo, A. (2022). Human performance in manufacturing tasks: Optimization and assessment of required workload and capabilities. *Safety Science*, *154*, 105838.

Lim, Yixiang, Alessandro Gardi, Roberto Sabatini, Subramanian Ramasamy, Trevor Kistan, Neta Ezer, Julian Vince, and Robert Bolia. 2018. “Avionics Human-Machine Interfaces and Interactions for Manned and Unmanned Aircraft.” Progress in Aerospace Sciences 102 (October): 1–46. https://doi.org/10.1016/j.paerosci.2018.05.002.

Liu, J. e. (2021). A study on assigning performance shaping factors of the SPAR-H method for adequacy human reliability analysis of nuclear power plants. . *International Journal of Industrial Ergonomics 81*, 103051.

Matheson, E., Minto, R., Zampieri, E. G. G., Faccio, M., & Rosati, G. (2019). Human-robot collaboration in manufacturing applications: A review. In *Robotics* (Vol. 8, Issue 4). MDPI AG. https://doi.org/10.3390/robotics8040100

Meteier, Quentin, Marine Capallera, Simon Ruffieux, Leonardo Angelini, Omar Abou Khaled, Elena Mugellini, Marino Widmer, and Andreas Sonderegger. 2021. “Classification of Drivers’ Workload Using Physiological Signals in Conditional Automation.” Frontiers in Psychology 12 (February): 268. https://doi.org/10.3389/FPSYG.2021.596038/BIBTEX.

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

Michalos, G., Makris, S., Tsarouchi, P., Guasch, T., Kontovrakis, D., & Chryssolouris, G. (2015). Design considerations for safe human-robot collaborative workplaces. *Procedia CIRP*, *37*, 248–253. https://doi.org/10.1016/j.procir.2015.08.014

Mijović, Pavle, Vanja Ković, Maarten De Vos, Ivan Mačužić, Petar Todorović, Branislav Jeremić, and Ivan Gligorijević. 2017. “Towards Continuous and Real-Time Attention Monitoring at Work:

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Reaction Time versus Brain Response.” Ergonomics. https://doi.org/10.1080/00140139.2016.1142121.

Mkrtchyan, L. L. (2015). Bayesian belief networks for human reliability analysis: A review of applications and gaps. *Reliability engineering & system safety 139*, 1-16.

Murdoch, B. (2021). Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Med Ethics 22, 122. Available online at: https://doi.org/10.1186/s12910-021-00687-3

N. Ratliff, D. Bradley, J. A. Bagnell, and J. Chestnutt. Boosting structured prediction for imitation learning. In Advances in Neural Information Processing Systems 19, 2006.

Niekum, Scott, Sachin Chitta, Andrew G. Barto, Bhaskara Marthi, and Sarah Osentoski. "Incremental Semantically Grounded Learning from Demonstration." In Robotics: Science and Systems, vol. 9, pp. 10-15607. 2013.

Osa, Takayuki, Joni Pajarinen, Gerhard Neumann, J. Andrew Bagnell, Pieter Abbeel, and Jan Peters. "An algorithmic perspective on imitation learning." *Foundations and Trends® in Robotics* 7, no. 1-2 (2018): 1-179.Park, J. A. (2019). A comparison of the quantification aspects of human reliability analysis methods in nuclear power plants. . *Annals of Nuclear Energy 133*, 297-312.

Peterson, Steven M., Rajesh P.N. Rao, and Bingni W. Brunton. 2022. “Learning Neural Decoders without Labels Using Multiple Data Streams.” Journal of Neural Engineering 19 (4). https://doi.org/10.1088/1741-2552/ac857c.

Podofillini, L. a. (2020). Foundations and novel domains for Human Reliability Analysis. *Reliability Engineering & System Safety 194*, 106759.

Ponce, A. (2021). The AI Regulation: Entering an AI Regulatory Winter? Why An Ad Hoc Directive on AI in Employment Is Required. SSRN Scholarly Paper No. 3873786. Available at: https://doi.org/10.2139/ssrn.3873786Poria, Soujanya, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. “A Review of Affective Computing: From Unimodal Analysis to Multimodal Fusion.” Information Fusion 37 (September): 98–125. https://doi.org/10.1016/j.inffus.2017.02.003.

Putze, Felix, and Tanja Schultz. 2014. “Adaptive Cognitive Technical Systems.” Journal of Neuroscience Methods 234 (August): 108–15. https://doi.org/10.1016/j.jneumeth.2014.06.029.

Rasmussen, J. K. (1987). New Technology and the Effect of Human Error. *Wiley*.

Reason, J. (2000). Human error: Models and management. In *British Medical Journal* (Vol. 320, Issue 7237). https://doi.org/10.1136/bmj.320.7237.768

Ren, Shan, Yingfeng Zhang, Yang Liu, Tomohiko Sakao, Donald Huisingh, and Cecilia M V B Almeida. 2019. “A Comprehensive Review of Big Data Analytics throughout Product Lifecycle to Support Sustainable Smart Manufacturing: A Framework, Challenges and Future Research Directions.” Journal of Cleaner Production 210: 1343–65. https://doi.org/https://doi.org/10.1016/j.jclepro.2018.11.025.

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Robbins, S. & van Wynsberghe, A. (2022). Our New Artificial Intelligence Infrastructure: Becoming Locked into an Unsustainable Future. Sustainability 14, 4829. Available at: doi.org/10.3390/su14084829Savković M, Caiazzo, C., Djapan, M., Vukićević, A. M., Pušica M, Mačužić, I. (2022). Development of Modular and Adaptive Laboratory Set-Up for Neuroergonomic and Human-Robot Interaction. *Research. Front. Neurorobot. 16:863637*.

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D., Glasstetter, M., Eggensperger, K., Tangerman, M., . . . Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping, 38*(11), 5391-5420.

Schoeffer et al. (2022). On Explanations, Fairness, and Appropriate Reliance in Human-AI Decision-Making. Human-Computer Interaction at Cornell University. Available at: https://arxiv.org/abs/2209.11812Sellen, A., Norman, D. A., Rasmussen, J., Duncan, K., & Leplat, J. (1989). New Technology and Human Error. *The American Journal of Psychology*, *102*(1). https://doi.org/10.2307/1423121

Segura Anaya L.H. et al. (2018). Ethical Implications of User Perceptions of Wearable Devices. Sci Eng Ethics ;24(1):1-28. doi: 10.1007/s11948-017-9872-8. PMID: 28155094.

Shoumy, Nusrat J., Li Minn Ang, Kah Phooi Seng, D. M.Motiur Rahaman, and Tanveer Zia. 2020. “Multimodal Big Data Affective Analytics: A Comprehensive Survey Using Text, Audio, Visual and Physiological Signals.” Journal of Network and Computer Applications 149 (January). https://doi.org/10.1016/J.JNCA.2019.102447.

Song, Tengfei, Wenming Zheng, Peng Song, and Zhen Cui. 2018. “EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks,” 1949–3045. https://doi.org/10.1109/TAFFC.2018.2817622.

Stancin, I., Cifrek, M., & Jovic, A. (2021). A review of EEG signal features and their application in driver drowsiness detection systems. *Sensors, 21*(11), 3786.

Statista Search Department (2015, September 15th) System costs of a typical spot-welding robot in the U.S. automotive industry from 2005 to 2025 [Infographic]. Statista. https://www.statista.com/statistics/561605/system-costs-of-a-spot-welding-robot-in-us-automotive-industry/

Tilmes, N. (2022). Disability, fairness, and algorithmic bias in AI recruitment. Ethics and Information Technology 24 (2). Available at: https://philpapers.org/rec/TILDFA-2

Todolí-Signes, A. (2019). Algorithms, artificial intelligence and automated decisions concerning workers and the risks of discrimination: The necessary collective governance of data protection. European Review of Labour and Research, 25(4), 465–481. https://doi.org/10.1177/1024258919876416

Todolí-Signes, A. (2020) Complying with the First Law of Robotics: An Analysis of the Occupational Risks Associated with Work Directed by an Algorithm/Artificial Intelligence (January 20, 2020).

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Tu, J. & Gao, W. (2021). Ethical Considerations of Wearable Technologies in Human Research. Advanced Healthcare Materials. DOI: 10.1002/adhm.202100127

Villani, Valeria, Fabio Pini, Francesco Leali, and Cristian Secchi. "Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications." Mechatronics 55 (2018): 248-266.

Vortmann, Lisa Marie, Simon Ceh, and Felix Putze. 2022. “Multimodal EEG and Eye Tracking Feature Fusion Approaches for Attention Classification in Hybrid BCIs.” Frontiers in Computer Science 4 (March): 31. https://doi.org/10.3389/fcomp.2022.780580.

Wei Liu, Jie-Lin Qiu, Wei-Long Zheng, and Bao-Liang Lu. 2021. “SEED Dataset.” 2021. https://bcmi.sjtu.edu.cn/home/seed/seed-v.html.

Wildhaber, I. (2018). Artificial intelligence and robotics, the workplace, and workplace-related law. Research Handbook on the Law of Artificial Intelligence, 577–608. ISBN: 978 1 78643 904 8

Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). Behind human error: Cognitive systems, computers and hindsight. CSERIAC Soar, December.

Wu, D., Sharma, N., & Blumenstein, M. (2017). Recent advances in video-based human action recognition using deep learning: A review. *Proceedings of the International Joint Conference on Neural Networks*, *2017-May*, 2865–2872. https://doi.org/10.1109/IJCNN.2017.7966210

Wu, Qun, Nilanjan Dey, Fuqian Shi, Rubén González Crespo, and R Simon Sherratt. 2021. “Emotion Classification on Eye-Tracking and Electroencephalograph Fused Signals Employing Deep Gradient Neural Networks.” Applied Soft Computing 110: 107752. https://doi.org/https://doi.org/10.1016/j.asoc.2021.107752.

Yang, Shuo, Zhong Yin, Yagang Wang, Wei Zhang, Yongxiong Wang, and Jianhua Zhang. 2019. “Assessing Cognitive Mental Workload via EEG Signals and an Ensemble Deep Learning Classifier Based on Denoising Autoencoders.” Computers in Biology and Medicine 109 (June): 159–70. https://doi.org/10.1016/J.COMPBIOMED.2019.04.034.

Yin, Yongqiang, Xiangwei Zheng, Bin Hu, Yuang Zhang, and Xinchun Cui. 2021. “EEG Emotion Recognition Using Fusion Model of Graph Convolutional Neural Networks and LSTM.” Applied Soft Computing 100 (March): 106954. https://doi.org/10.1016/J.ASOC.2020.106954.

Yu, Baek Woon, Ji Hoon Jeong, Dae Hyeok Lee, and Seong Whan Lee. 2020. “Detection of Pilot’s Drowsiness Based on Multimodal Convolutional Bidirectional LSTM Network.” Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12047 LNCS: 530–43. https://doi.org/10.1007/978-3-030-41299-9_41.

Yufeng, K., Tao, J., Liu, S., Yong, C., Xuejun, J., Jin, J., & Dong, M. (2021). Cross-Task Consistency of Electroencephalography-Based Mental Workload Indicators: Comparisons Between Power Spectral Density and Task-Irrelevant Auditory Event-Related Potentials. *Frontiers in Neuroscience, 15*.

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901

Zhang, Pengbo, Xue Wang, Junfeng Chen, Wei You, and Weihang Zhang. 2019. “Spectral and Temporal Feature Learning with Two-Stream Neural Networks for Mental Workload Assessment.” IEEE Transactions on Neural Systems and Rehabilitation Engineering 27 (6): 1149–59. https://doi.org/10.1109/TNSRE.2019.2913400.

Zhang, Ting, Jing Yang, Nade Liang, Brandon J. Pitts, Kwaku O. Prakah-Asante, Reates Curry, Bradley S. Duerstock, Juan P. Wachs, and Denny Yu. 2020. “Physiological Measurements of Situation Awareness: A Systematic Review.” Human Factors, November. https://doi.org/10.1177/0018720820969071/ASSET/IMAGES/LARGE/10.1177_0018720820969071-FIG1.JPEG.

Zhong, Peixiang, Di Wang, Senior Member, and Chunyan Miao. 2019. “EEG-Based Emotion Recognition Using Regularized Graph Neural Networks.” IEEE Transactions on Affective Computing, July, 1–1. https://doi.org/10.1109/taffc.2020.2994159.

This project has received funding from the European Union’s Horizon 2020- MSCA-ITN-2020 under grant agreement No. 955901