AI REGIO

D7.1 Legal, Regulatory and Ethical Framework

Author:	ETA
Work Package:	WP7
Delivery date:	02.11.2021
Due date:	30.09.2021
Classification:	PU
Type:	Report

H2020 Innovation Action - This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N. 952003

Status of deliverable

Action/role	Name	Date (dd.mm.yyyy)
Submitted by	Marina Cugurra (ETA)	02.11.2021
Responsible (WP leader)	Pieter Vernooij (BPI)	27.10.2021
Approved by (internal reviewer)	Angelo Marguglio (ENG)	22.10.2021

Revision History

Date (dd.mm.yyyy)	Version	Author	Comments
25.01.2021	v0.1	ETA	ToC initial draft
28.06.2021	v0.2	ETA	ToC final draft considering partners' feedback and project's progress
19.07.21	v0.3	ETA, POLIMI	First round contributions
30.07.2021	v0.4	ETA, BPI	Second round contribution
06.08.2021	v0.5	ETA	First draft of the document
03.09.2021	v0.6	ETA, BPI	Integration/ refinements of previous contributions
22.09.2021	v0.7	STIIMA, ETA	Third round contributions
29.09.2021	v0.8	ETA	Second draft of the document
15.10.2021	v0.9	STIIMA, SKU, ETA	Fourth round contribution and Integration/ refinements of previous contributions
19.10.2021	v0.10	ETA	Document ready for Peer Review and Quality Check
22.10.2021	v0.11	ENG	Quality Check and Peer Review
27.10.2021	v0.12	BPI	WP7 Leader's approval
02.11.2021	v1.0	ETA, POLIMI	Final release and submission to the EC

Author(s) contact information

Name	Organisation	E-mail
Marina Cugurra	ETA	marina.cugurra@eta-one.com
Gianfranco Modoni	STIIMA	Gianfranco.Modoni@stiima.cnr.it
Pieter Vernooij	BPI	p.vernooij@brainportindustries.nl
Tibor de Kroon	BPI	t.dekroon@brainportindustries.nl
Sergio Gusmeroli	POLIMI	sergio.gusmeroli@polimi.it
Francisco Souza	SKU	francisco.souza@ru.nl

Contents

Figures

Abbreviations and Acronyms:	
AI	Artificial Intelligence
AI Act	Artificial Intelligence Act (proposal)
AIDA	Special Committee on the Artificial Intelligence in the Digital Age
ALTAI	Assessment List for Trustworthy Artificial Intelligence
API	Application Programming Interface
B2B	Business to Business
CA	Competent Authority
CI	Collaborative Intelligence
DF	Didactic Factory
DIH	Digital Innovation Hub
DMA	Digital Market Act (proposal)
DMP	Data Management Plan
DOA	Description of Actions
DPO	Data Protection Officer
DR BEST	Data, Remote, Business, Ecosystem, Skills, Technology
DSA	Digital Service Act (proposal)
DT	Digital Twin
EC	European Commission

EDIH	European Digital Innovation Hub
EEG	Electroencephalogram
EFFRA	European Factories of the Future Research Association
ELS	Ethical, Legal and Societal
EP	European Parliament
ePD	e-Privacy Directive
ePR	e-Privacy Regulation (proposal)
ES	Ethical Strategy
EU	European
GDPR	General Data Protection Regulation (Regulation EU 2016/679)
HF	Human Factor(s)
HLEG	High Level Expert Group
HMI	Human-Machine Interaction
HRL	Human Rights Law
IAs	Industry Agreements
IDS	International Data Space
IEEE	Institute of Electrical and Electronics Engineers
IP	Intellectual Property
IPR	Intellectual Property Rights
KHC	Know-How-to-Cooperate
KPI	Key Performance Indicator
OEM	Original Equipment Manufacturer
MSD	Musculoskeletal Disorders
OSAI	Observatory on Society and Artificial Intelligence
PIA	Privacy Impact Assessment
R	Remotization
RED	Radio-Equipment Directive
SB	Standardisation Body
SME	Small Medium Enterprise
TEF	Testing and Experimentation Facility
TERESA	Technological and Regulatory Sandbox
UI	User Interface
VI	Vanguard Initiative
WP	Work Package

1 Executive Summary

This document is aimed at developing a threefold trustworthy framework for the project, rotating around AI REGIO TERESAs.

The first pillar of the framework consists of the AI REGIO Legal, Regulatory and Ethical Framework, where key topics, like Liability and Data Ownership and related challenges are tackled. Moving ahead from the identification and description of the relevant sources as outlined in D2.7 "Legal and Ethical Requirements and Guidelines", key challenges have been deepened and they might be potentially further investigated and tested in the TERESAs.

The second pillar lingers over the Human Factor concept, role and implications within AI REGIO Collaborative Intelligence Model and the foreseen shift towards Industry 5.0, with AI-driven autonomous systems efficiently and effectively interacting with Humans according to such Model. This entails dealing with the human-centric aspects of AI-based manufacturing systems, as well as enabling an immersive, value-driven AI-based digital workplace. The fruitful humans-machines collaboration brought by CI breakthroughs comprises a bidimensional dimension, both technical and ethical, legal and societal (ELS). The document depicts both the aspects and their relevance in AI REGIO, in order to ensure that technical and ELS implications are consistent and proceed in a coherent and harmonized manner.

The last, key pillar is the AI REGIO TERESA Framework. It describes the concept and methodology adopted in the project and depict the roles to be played, respectively, by regional Didactic Factories and by the Competent Authorities in this exercise. The document also includes the template to be used for collecting and reporting information from AI REGIO TERESA. Such template consists of two parts: an overview of the TERESA concerned lingering over one or more selected topics (legal at ethical issues at stake, Human Factor issues at stake) and the testing plan. The same format will be common to each of the TERESAs.

The deliverable also provides the roadmap and next steps that will be followed for TERESAs' deployment in the next phase of the project, when they will be effectively implemented. It is completed by the conclusions.

2 Introduction

2.1 About this deliverable

This document pertains to WP7.1 "AI DIH Legal, Regulatory and Ethical Issues", aimed at developing a threefold Trustworthy Framework for AI REGIO. The framework is directed to tackle with privacy issues and to move forward towards a clear liability system and to uphold EU ethical values, as well as to set and implement live, time-bound testing of AI REGIO innovative solutions in technology-regulatory sandboxes, and to investigate the key features of the Human Factor dimension within AI REGIO Collaborative Intelligence (CI) Model, focusing on the bidimensional nature, technical and ethical/regulatory. In this context, this deliverable is directed to set such Trustworthy Framework for the project, rotating around AI REGIO TERESAs, notably by:

- deepening the key legal and ethical challenges for AI REGIO and its CI Model, starting from the key pieces of legislation identified in D2.1 as part of the Legal, Regulatory and Ethical Framework relevant to AI REGIO development and future uptake. These challenges include, for instance, Privacy and Data Protection, Liability, Data Ownership and Sovereignty and other ethical concerns;
- depicting the Human Factor concept, role and implications within AI REGIO Collaborative Intelligence Model and the foreseen shift towards Industry 5.0.The human-centric CI Approach drives the whole activities related to HFs in the project, with specific focus both on the human-to-machine element and to the machine-to-human one. The investigation on the bidimensional nature, technical and ethical/regulatory, of HF and CI paradigm in AI REGIO is also essential, in this perspective;
- lingering over the TERESA concept and methodology and respective roles to be played by the regional Didactic Factories and Competent Authorities in this exercise, including the model form to be used to plan and execute TERESA's operations. The plans for next period will be presented for such deployment.

The document and underlying activities are mainly interrelated with:

- WP2 "Beyond REQUIREMENTS: AI DIH Digital Transformation from scenarios to business cases", in particular WP2.4 "Legal and Ethical Requirements Specification in Collaborative Intelligence Scenarios", where the regulatory and ethical framework relevant to AI REGIO has been analysed (and will be further analysed, if necessary) in a systematic way to elicit the legal and ethical requirements reported in order to deliver a value-driven and legal-respectful technology. The main regulatory sources mentioned in it are going to be described in this deliverable, with a possible integration and update in its next release at M24 /D7.2), taking into account both the project progresses and the new pieces of legislation (both applicable and under development);
- WP4 "Beyond PLATFORMS: AI DIH Open Platforms and DIH platform" and WP5 "Beyond INDUSTRY 4.0: AI DIH Industry 5.0 and Data Sharing Spaces", since the regulatory sources deepened in this deliverable and the related regulatory and ethical challenges are relevant for the design, development and/or customisation of key technological assets and bundles, but also of overall AI REGIO system, and, as mentioned, drove the elicitation of the legal and ethical requirements, as reported in D2.1 "Legal and Ethical Requirements and Guidelines";

- WP6 "Beyond EXPERIMENTS: AI DIH regional facilities from demos to champions", since most of the challenges identified and remarks resulting from the analysis of the legislation might be relevant to many experiments;
- WP1 "Beyond COORDINATION: AI DIH consortium from management to governance", besides for what concerns the AI REGIO Ethical Policy set in D1.11 "Ethics Assessment and Data Management Plan", also for the two waves of open calls for third parties to complement the achievements of the project and its experimentations, to be conducted under WP1.3 "Innovation Coordination, Business Impact and Open Calls Management": one or more of the additional experiments that will be selected might include a TERESA where to test and examine also the regulatory and ethical implications, for instance on human-machine interaction.

2.2 Document structure

The document is structured as follows:

- **Section 3** provides an analysis of the main regulatory and ethical challenges for AI REGIO socio-technical system in the evolving legal landscape, focusing on liability and safety, data ownership and sovereignty, privacy and data protection, as well as on other ethical issues;
- **Section 4** contains the finding of an in-depth investigation of the Human Factor within AI REGIO Collaborative Intelligence Model, focusing on the Human-centric CI Approach in the project in terms of "humans in the loop", paying attention both to the technical and the ELS dimension of HF in AI REGIO. The section also includes insights on the Human Factors and ergonomics management, addressed as an aspect of the CI interactions;
- **Section 5** depicts the AI REGIO TERESA Framework, including the concept of TERESA and the methodology driving its conception and implementation, paying attention to the Didactic Factories definition and role in the TERESA development, as well as to the concept and role of the Competent Authorities in the same context. The section also comprises the plans for the next period;
- **Section 6** draws conclusions.
- **Annex 1** contains the form to be filled in by each of the Didactic Factories hosting one TERESA and includes an overview of the experiments and a detailed testing plan.

3 Regulatory and ethical challenges for AI REGIO socio-technical system in the evolving legal landscape

3.1 Liability and Safety

In the collaborative manufacturing supply chains challenges regarding the allocation of the duties and liability are brough by the "servitization" of the manufacturing process, which softens the distinction between the phases of "manufacturing" and "operation" and where new figures are involved playing different roles in an evolving ecosystem of actors and cyber-physical artefacts[1] [1], as well as AI solutions, like in the AI REGIO project. From a liability perspective, the potential culpability for anything that could go wrong might quickly shift from one component to another: the identification of who is responsible for what when something has gone wrong in a given situation can

[1] Gusmeroli S., Dalle Carbonare D, «BDVA White Paper "Big Data challenges in Smart Manufacturing Industry",» 2020.

be layered when many actors come into play, including AI developers, algorithm trainers, data collectors, controllers, and processors, as well as manufacturers of the devices incorporating the AI software and owners of the software (which might not coincide with the developers) and the final users of the devices.

This implies the question of the identification of "who controls whom" and "who controls what" and, thereby, who is accountable and for what. The allocation of liability is even more complex when the human behaviour is mediated through an autonomous system, like in the AI REGIO context.

The AI-empowered industrial procedures, programs and collaborative tools in the "collaborative manufacturing supply chains" relies on the value of preserving a log of the different actions taking place in a procedure, for which it is key to collect and demonstrate a historic of actions liability in order to spot the damage or violations and treat them with ease, making the whole collaborative experience efficient, whilst avoiding misunderstandings potentially occurring when working and supervising autonomous systems.

Liability is a key aspect for the companies when dealing with new services and solutions like those of AI REGIO and plays a significant role in the digital transformation: the identification of liable and reliable stakeholders or partners is critical and it is key to have clear understanding of responsibilities between different actors also when an AI system used in the productive process makes mistakes producing damages or injury to property and human beings.

Potential harms might include, for instance, unavoidable or inherent harms, deliberate or leastcost arms, defect-driven harms, misuse harms, unforeseen harms, systemic harms, as well as collateral harms.

When a harm occurs, questions of attribution and remedies will arise, notably regarding the attribution for AI-induced harms and the identification adequate mechanisms to mitigate possible AI harms. Some example of these questions are: "Whose fault is it if an AI system takes a decision which causes harm?", "How to identify and apportion such a fault?", "What sort of remedies should be imposed or measures should be taken to avoid the repetition of such mistakes in the future?"

In order to answer these questions it is necessary to examine the **intersection of products liability and AI**.

The White Paper on Artificial Intelligence – A European Approach to excellence and Trust[2] [2] and the Report on safety and liability[3] [3] underline that, **in order to strengthen the AI growth and its wide uptake, the liability topic should be properly taken tackled** at policy level, especially the **liability for damage caused by AI-systems**, and efforts should be directed to address the uncertainty regarding the allocation of responsibilities between different actors. In addition, the Resolutions adopted by the European Parliament in October 2020[4] [4] [5], covering ethics and civil liability, calls for the harmonization of the legal framework for civil liability claims and for a regime of strict liability on operators of high-risk AI systems.

Currently the legal framework is characterized by the partially harmonised EU legislation on liability for defective products (**Directive 85/374/EEC on liability for defective products**), applicable to any product marketed in the European Economic Area. It mainly states that:

i) if a defective product causes any physical damage to consumers or their property, compensation has to be provided by the producer, even when there is no negligence or fault on their part (the so-called "**strict liability regime**"). Under strict liability, reflecting the view

[2] European Commission, «COM (2020) 65 final "White Paper on Artificial Intelligence – A European Approach to Excellence and Trust,» 2020.

[3] European Commission, «COM(2020) 64 final. "Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics",» 2020

[4] European Parliament, «Resolution on a civil liability regime for artificial intelligence, (2020/2014 (INL),» 2020 and European Parliament, "Resolution on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012 (INL)", 2020

that consumers have a right to expect safe products, the manufacturers (including those producing AI products) are held liable for unsafe defects without requiring an inquiry as to whether the defect arose from an identifiable failure, like a design defect, a manufacturing defect, or manufacturer negligence. The consumer who suffers a resulting harm would not have the burden of identifying specifically where in the design or manufacturing process the defect was introduced;

ii) **the producers can be cleared of liability only under certain conditions**, for instance if they prove that the state of scientific or technical knowledge at the time the product was put into circulation could not detect the defect. The manufacturers' obligation is to make products that will be safe when used in reasonably foreseeable ways.

On the other hand, as highlighted by the EC's Report on the safety and liability implications of Artificial Intelligence[5] [3], the Internet of Things and robotics, **the framework at national level is fragmented** and there are not yet clear liability rules specifically applicable to damage resulting from the use of emerging digital technologies such as AI (with the limited exception of highly or fully automated vehicles).

The harmful effects of AI (as those caused by the operation of other emerging digital technologies) **can be compensated under the tort law of EU Member States**, which is **largely non-harmonised** (with the exception of product liability law under Directive 85/374/EC): the significant differences between the national regimes may not always lead to satisfactory results, giving rise to different outcomes depending on which jurisdiction is applicable.

The AI systems have the key characteristic to be able to learn, or, stated another way, to go beyond the simple implementation of human-designed algorithms, instead, creating their own algorithm (sometimes by revising algorithms originally created by humans, and sometimes completely from scratch). This might raise complex issues in relation to products liability, in particular to understand whether companies need to bear responsibility for the AI products they create, even when those products evolve in ways not specifically desired or foreseeable by their manufacturers, as well as to apportion blame and responsibilities when there are multiple companies that have had a hand in designing an AI system (or in shaping the post-sale algorithm evolution).

Some scholars[6] [6] argue that **possible defenses which could be asserted by the AI company targeted by a products liability lawsuit** range, different from the not legitimate "it's the algorithm's fault", range from:

- **Blaming the AI**, in particular the AI-driven evolution of the algorithms that the company originally designed: the company could mention the fact that the AI algorithm has evolved in a way that introduced defects. However, this usually will not let company escaping liability, since if they want to reap the benefits of intelligent algorithms, they also need to accept the attendant risks. It is known at the time of the original sale that AI enables learning and automated post-sale changes to the algorithm aimed at improving its performance (though the company would not know specifically how the AI algorithm might evolve): this is usually portrayed as an asset to prospective customers (as a benefit) and reflected in the product pricing or in the marketing strategies. Therefore, if it turns out that the evolution renders the product harmful, the company needs to bear responsibility for that as well.
- **Blaming the data**, motivating that the problems were due to the low quality of the data provided to the system that was used as a basis for the AI-driven algorithm evolution. In other words, companies might assert that there is nothing wrong with the AI system and the bad data caused the AI evolution in harmful ways. Except maybe in some limited cases, such as

[5] European Commission, «COM(2020) 64 final. "Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics",» 2020

[6] John Villasenor, "Products Liability law as a way to address AI harms", report from The Brookings Institution's Artificial Intelligence and Emerging Technology (AIET) Initiative, AI Governance series, 2019.

where a malicious user intentionally provides a system with data intended to sabotage its performance, this "blame the data" argument is not suitable, since the producers of the AI system are responsible to anticipate the types of data that might be provided under reasonably foreseeable usage scenarios and that the resulting algorithm evolution is beneficial or neutral, not detrimental;

- **Blaming the users,** in particular the operations of the human users of the AI system. Users of an AI-based system might bear partial or full responsibility for harms arising in association with such system, in case applies it in clearly inappropriate ways. However, the users can not be credibly blamed if they engaged with an AI system in reasonably foreseeable ways and, in doing so, they inadvertently cause it to evolve in a manner introducing defects producing immediate or later harm. The human-AI interface, which raises its own set of products liability issues, needs to be adequate and take into account the interactions among the different software components of the AI system with the human beings;
- **Blaming the upstream or downstream supply chain**, which means other companies in the supply chain either downstream or upstream from the company being sued. Typically, there are multiple suppliers upstream of a consumer, such as the company that sold the product directly to the consumer, which might have purchased a software component of the product from another entity. This entity, in turn, might have built some parts of the software in-house and licensed other portions from yet another company. The allocation of blame within the supply chain implies, besides the technical analysis regarding the sources of various aspects of the AI algorithm, also the consideration of the legal agreements among the companies involved (including on indemnification aspects).

On the other hand, harm caused by an AI system could be the direct consequence of its programming or of its **negligent design, training, or operation** (e.g., lack of adequate cybersecurity protections), as well as of an unforeseeable harm generated by an interaction with unforeseeable real-world data.

The **liability risk associated with AI systems might differ depending on the nature of the AI** (for instance, whether the AI system is sold or licensed as a software or service or it is embedded in a tangible device), as well as on the function of the AI output: predictive systems, for instance, differ from fully autonomous systems where humans seems to be "out of the loop", thought under European Union law, any automated decision having legal or similar effects on individuals, as AI decisions may, needs to be subject to some type of **human oversight**. In the latter case, the liability question is even more complex and another human being (or perhaps his or her legal person principal) is put at risk of liability.

Some authors wonder whether AI systems merit a new approach to liability, such as giving **AI Legal Personhood** and it is opportune to held them liable, especially in case of complex AI Ecosystem where it is hard to pin onto a particular tortfeasor (or group of them). This idea of granting AI systems personhood is a legal fiction and rests on premises that AI system should be capable of holding assets either directly (similarly to a corporation) or indirectly (assuming that either the licensor or the licensee of the AI system should act on the AI system's behalf). However, **this hypothesis was refused by the European Parliament**. In the previously mentioned three resolutions on AI covering ethics, civil liability, and intellectual property (IP) published on October 2020 it calls for a better harmonisation of the legal framework for civil liability claims and for a regime of strict liability on operators of high-risk AI systems. In the resolution on civil liability it is clarified that "all physical or virtual activities, devices or processes that are driven by AI systems may technically be the direct or indirect cause of harm or damage, yet are nearly always the result of someone building, deploying or interfering with the systems… in this respect that it is not necessary to give legal personality to AI-systems".

The European Parliament also set up a **Special Committee on the Artificial Intelligence in the Digital Age (AIDA)**, in order to provide a "holistic approach providing a common, long-term position that highlights the European Union's key values and objectives relating to artificial intelligence in the digital age", as well as to ensure that the digital transformation path is in line with **human rights' protectio**n and is human-centric.

The Resolution on "Civil liability regime for artificial intelligence"[7] [6] acknowledges that "the challenge related to the introduction of AI-systems into society, the workplace and the economy is one of the most important questions on the current political agenda… technologies based on AI could and should endeavour to improve our lives in almost every sector, from the personal sphere.. to the working environment, for example alleviation from tedious and repetitive tasks". The EU Parliament states that, despite there is no need for a complete revision of the liability regimes, "the complexity, connectivity, opacity, vulnerability, the capacity of being modified through updates, the capacity for self-learning and the potential autonomy of AI systems, as well as the multitude of actors involved represent nevertheless a significant challenge to the effectiveness of Union and national liability framework provisions" and therefore **specific and coordinated adjustments to the "liability regimes are necessary** to avoid a situation in which persons who suffer harm or whose property is damaged end up without compensation".

The EC, with the support of the large majority of the stakeholders from the public and private sectors in a public consultation run between February and June 2020, is also setting up a European regulatory framework for trustworthy AI, as well as the **partial revision of the existing Product Liability Directive** to cover particular risks engendered by certain AI applications. As part of this framework, the EC published on 20 April 2021 its **Proposal for a Regulation on a European approach for Artificial Intelligence**[8] [8] and on 30 June 2021 also published an **Inception Impact Assessment**[9] [9] on the Initiative on a likely legislative initiative to adapt the EU liability rules to the digital age and circular economy (expected at the end of 2021 or at the beginning of 2022).

The AI Act also addresses the issue of civil liability for AI systems, imposing specific obligations upon providers, importers, users, distributors, and even third parties (Articles 16 to 29), thereby validating the approach established in the October 2020 (in the Resolution of the European Parliament on the civil liability regime for AI), based on the assumption that "AI-systems have neither legal personality nor human conscience". The Recital 53 of the AI Act sets that "it is appropriate that a specific natural or legal person, defined as **the provider, takes the responsibility for the placing on the market or putting into service of a high-risk AI system**, regardless of whether that natural or legal person is the person who designed or developed the system."

Therefore the AI Act opts for **a risk-based approach**. On this regard, the Report on the civil liability regime for AI, which follows the approach proposed by the White Paper to regulate "high-risk" AI applications involving significant risks both in the sector and in its intended use (especially from a safety, consumer rights and fundamental rights perspective), states that **the operators of "high-risk" AI-systems would be subject to strict liability for any damage** that results in harm to life, health, damage to property or harm that results in economic loss. This means that operators of high-risk AI-systems will be liable for any harm caused by an autonomous activity, device or process driven by their AI system, even if they did not act negligently (**strict liability regime**). In **situations where there is more than one operator**, all operators should be jointly and severally liable, and have the right to recourse proportionately against each other.

[7] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL))

[8] European Commission, «COM/2021/206 final, Proposal for a Regulation laying down harmonized rules on Artificial Intelligence (ARTIFICIAL INTELLIGENCE ACT),» 2021.

[9] European Commission, Inception Impact Assessment on the Initiative "Adapting liability rules to the digital age and circular economy", 2021.

The AI Act defines the "**Operator**" as "both the frontend and the backend operator as long as the latter's liability is not already covered by the Product Liability Directive". The **frontend operator** is "any natural or legal person who exercises a degree of control over a risk connected with the operation and functioning of the AI-system", whilst the **backend operator** is "the natural or legal person who, on a continuous basis, defines the features of the technology, provides data and essential backend support service and therefore also exercises a degree of control over the risk connected with the operation and functioning of the AI-system".

The operator can exercise a certain level of control over the risk that the item poses: therefore any operator's action might affect the manner of the operation from the beginning to the end, by determining the input, output or results, or could change specific functions or processes within the AI-system. In case the frontend operator is also the producer of the AI system, the **AI Act** will prevail over the **Product Liability Directive**, whilst if the backend operator also qualifies as a producer under such Directive, then it will take precedence.

Moving to the **meaning of "high-risk" AI-system under the AI Act**, it means "a significant potential in an autonomously operating AI-system to cause harm or damage to one or more persons in a manner that is random and goes beyond what can reasonably be expected; the significance of the potential depends on the interplay between the severity of possible harm or damage, the degree of autonomy of decision-making, the likelihood that the risk materialises and the manner and the context in which the AI-system is being used".

To determine whether an AI-system is high-risk, the Report suggests to take into account also the sector in which significant risks could arise and the nature of the activities to be undertaken.

Whilst **most AI systems pose limited to no risk**, certain of them create risks that need to be addressed to avoid undesirable outcomes, such as unfairly disadvantage to certain individuals. Due to the insufficient protection offered by existing legislation to address the specific challenges AI systems may bring, the AI Act therefore provides rules that address the risks specifically created by AI systems, with a list of high-risk applications, sets **clear requirements for AI systems for high risk applications** and defines **specific obligations for AI users** and providers of high risk applications, besides a **conformity assessment** before the AI system is put into service or placed on the market, as well as enforcement after such an AI system is placed in the market and a governance structure at European and national level.

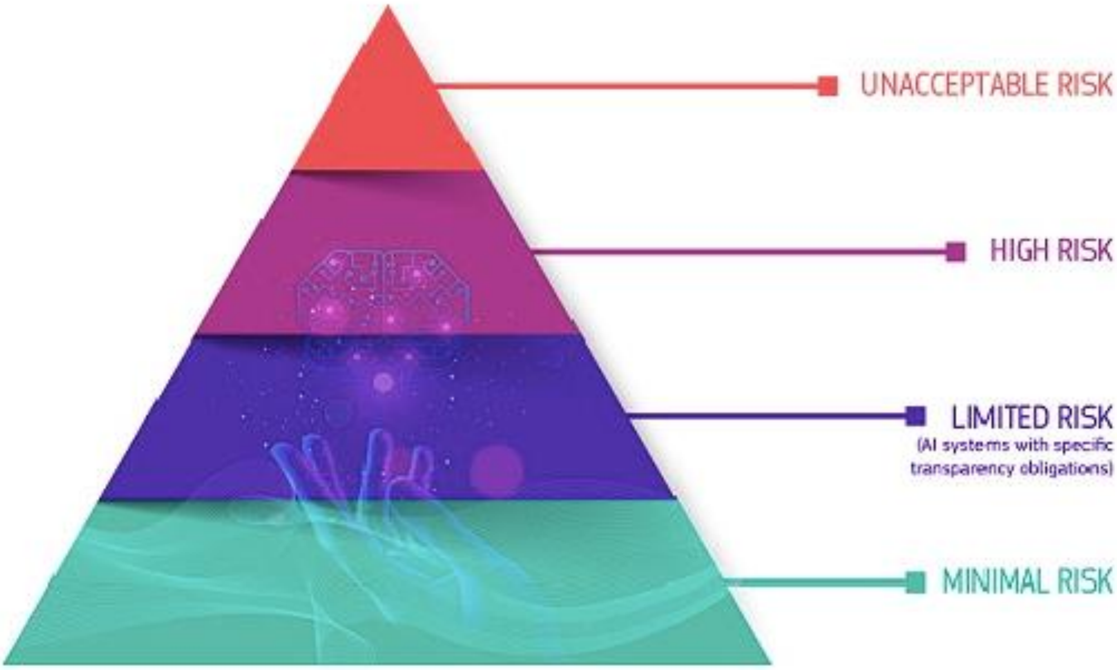


Figure 1. Classification of risk (source: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai)

As showed by figure 1, the **classification of the risks according to the AI Act** is as follows[10] [9]:

I. **Unacceptable risk**

This category includes the AI systems considered a clear threat to the safety, livelihoods and rights of people and they will be banned.

II. **High-risk**

This category includes all remote biometric identification and AI technology used, among others, in:

i) critical infrastructures (e.g. transport), in case they could put the life and health of citizens at risk;
ii) educational or vocational training, in case such system may determine the access to education and professional course of someone's life (e.g. scoring of exams);
iii) safety components of products (e.g. AI application in robot-assisted surgery);
iv) employment, workers management and access to self-employment (e.g. CV-sorting software for recruitment procedures);
v) essential private and public services (e.g. credit scoring denying citizens opportunity to obtain a loan);

This kind of AI systems will be subject to strict obligations before they can be put on the market, including i) adequate risk assessment and mitigation systems; ii) high quality of the datasets feeding the system in order to minimise risks and discriminatory outcomes; iii) logging of activity to ensure traceability of results iv) detailed documentation providing all information necessary for authorities to assess the compliance of the AI system; v) clear and adequate information to the user; vi) adequate human oversight measures to minimise risk; vii) high level of robustness, security and accuracy.

According to the AI Act, the operator of a "high-risk" AI-system "shall be strictly liable for any harm or damage that was caused by a physical or virtual activity, device or process driven by that AI-system", without being able to exonerate himself or herself by arguing that he or she acted with due diligence or that harm or damage was caused by an autonomous activity, device or process driven by the AI-system, except in case of force majeure (in this case the operator shall not be held liable for the harm or damage).

III. **Limited risk**

This category includes AI systems such as chatbots: in this case there are specific transparency obligations, so that the users, could be aware that they are interacting with a machine and can take an informed decision to continue or step back.

IV. **Minimal risk**

This category refers to applications such as AI-enabled video games or spam filters: for this kind of systems, which represents the vast majority of AI systems currently used in the EU, it is foreseen the free use since there is minimal or no risk.

It is expected that the **AI Act could enter into force in the second half of 2022** in a transitional period, during which standards would be mandated and developed. It could become applicable to operators during the second half of 2024, with the standards ready. In the next phase of the project it will be explored which kind of risk level imply AI REGIO tools.

The Inception Impact Assessment mentioned before[11] [9] clarifies that the overall objective of the **EU safety framework** "is to ensure that all products and services, including those integrating

[10] https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[11] European Commission, Inception Impact Assessment on the Initiative "Adapting liability rules to the digital age and circular economy", 2021

emerging digital technologies like robotics, Internet of things (IoT) and AI, operate safely, reliably and consistently", whilst "the **liability framework** provides for remedies if damage nevertheless occurs", (i) providing legal certainty to industry about the risk they take in the course of their business, (ii) encouraging the prevention of damage and (iii) ensuring injured parties are compensated.

The **existing liability framework** comprises:

- the Product Liability Directive 85/374/EEC, which harmonises at EU level the claims against the producer for damage caused to a consumer due to the defectiveness of a product. The producer is strictly liable for damage caused by a defect in their product, provided that the injured party proves the damage, the defect and the causal link between the two. The Directive applies to a vast range of products, including complex AI-driven devices.
- National liability regimes, which regulate several liability claims for damages caused by products and services: part of the claims is based on a liable person's conduct ('fault-based liability'), such as a producer, service provider or individual user of a product, whilst for others claims, a person identified by law (usually the operator, user or owner) is held liable independently of fault ('strict liability').

As regards product safety challenges, the Communication on "Building Trust in Human-Centric Artificial Intelligence"[12] [11] states that **AI systems and application should integrate safety and security-by-design mechanisms** to ensure the verifiability of their safety at every step, taking at heart the physical and mental safety of all individuals and groups potentially affected.

The existing product safety framework, still not fully harmonized and partially fragmented, mainly consists of the **General Product Safety Directive**, requiring that all consumer products (even if not regulated by the Union sectorial legislation) need to be safe. In addition, other pieces of legislation are relevant, including the **sectorial safety legislation** such as the Machinery Directive, the Radio-Equipment Directive (RED), the Medical Device Directive and the Toy Safety Directive, as well as the **standardization initiatives**.

In an effort directed to ensure consistency and complementary to present and future initiatives dwelling upon the same problems, **the EC promoted the revision of sectoral product legislation** (e.g., the Machinery Directive, the General Product Safety Directive) and other initiatives dedicated to the liability issues related to new technologies, including AI systems. For instance, on the same day that the AI Act was published, the EC also adopted the Proposal for a Regulation of the European Parliament and of the Council on machinery products, which will replace Directive 2006/42/EC on machinery.

High levels of safety for products and systems integrating new digital technologies and a liability framework with robust mechanisms remedying occurred damage are critical to a better protection of the consumers, which, in turn, creates **trust in AI technologies**, a prerequisite for their uptake by industry and users. This is expected to leverage the **competitiveness of EU Industry**.

3.2 Data ownership and sovereignty

In order to promote the well-functioning and competitive Digital Single Market with full deployment of the European Data Economy with relevant Business to Business (B2B) data sharing and re-use for the benefit of European Manufacturing sector, it is of utmost importance to identify and **remove the critical barriers and uncertainties to the development of the data economy** and the use of IoT, robots and autonomous systems. These barriers and uncertainties which might act as deterrents to companies want to enter the market or capture further segments. Among these barriers, which might affect the single company in different ways depending on its position in the value chain, its size and the sector in it operates, there are the **regulatory and ethical challenges related to**

[12] European Commission, COM(2019)168), "Building Trust in Human Centric Artificial Intelligence", 2019

share, access or (re-)use third party data and the data ownership, representing an impediment to data sharing.

In this regard, from a general viewpoint **three main categories of stakeholder involved in the Data Economy** come at stake:

- **Actors co-producing data** (product/service providers and product/service users). They have a different degree of control over the data: usually, the product/service provider retains the greatest degree of control over the data, whilst the user has more limited control, though, in some cases, the user retains greater control. The debate on 'data ownership' mainly affects these two types of actors, which are also the most relevant "data sharers", for different reasons.
- **Actors interested in accessing data** (providers' competitors and same sector down-stream providers).They are economic players (most often in the same value chain) that need the data for their business in order to ensure a level playing field for competition andbusiness mod-el innovation. This category includes competitors of the service producer and players downstream or upstream in the same value chain, despite they do not participate in the production of data. Therefore, for this category, which is the one suffering most from lack of access to data as their business model depends on the availability of them, it is not the ownership itself to come at stake, but rather the issue of access to data and the terms and conditions of access.
- **Actors interested in re-using data despite outside the sector** (data analytics companies and re-users of public interest data, but also universities, statistical offices etc.). For instance, it was detected in the manufacturing sector (but also in others) that the lack of data aggregation from many sources suffered by data analytics companies having access to the data of their clients without being able to aggregate it and (re-)use it, is detrimental to development of innovative artificial intelligence solutions, for which data aggregation is a prerequisite. Furthermore, data scientists could make use of data held by private players for reasons of public interest and for tackling with societal challenges.

With the shift of vision of data, from being a by-product of industrial, commercial, consumer and other activities to being a resource in their own right, their commercial value and importance grew and they are nowdays seen as the "**new oil**", thought, unlike oil, the data might be used by multiple actors and for multiple purposes. In this context, the data ownership issues have taken on great significance.

Regarding the **EU manufacturing industry** and the Industry 4.0 movement, they require the digitalization to face the increasingly competitive global landscape and data exchange within factories and in the wider manufacturing ecosystems, characterized by an evolving trend with decentralized and distributed data ecosystems associated with emerging business models and networks. The data are expected to improve the manufacturing companies' operations and strategic positioning, reducing costs, enhancing quality and at the same time flexibility and agility necessary to adapt to changes and to the needs of the customer, with new opportunities in several areas, such as: i) collaboration with customers, suppliers and service providers in order to increase supply chain visibility, work together on design, engineering, manufacturing, support and logistic challenges; ii) modular, more flexible and adaptive; iii) improvement of quality and lowering of maintenance costs through collecting, aggregating and analyzing data.

The **debate on data ownership**, which emerged early in the debate on access and (re-)use of data, is largely of major concern to the product/services providers and product/service users and only indirectly to all the other players in the value chain: in other words, in practical terms, the question of ownership directly concerns **two of the players in the value chain** (service or product provider and the user of the service), though it is essential the ability to have the right to access and use the data for specific purposes with sufficient clarity. In most use cases, the ownership of data automatically remains with the service/product providers of the data themselves.

The impact of "data ownership" claims is also strongly affected by the **distinction among the different types of data**, derived from the broad concept of "data", including also more abstract concepts such as the underlying information and derived knowledge: these can be classified based on the characteristics of its content (e.g. personal, non-personal, health data, machine data, etc.), as well as in relation to the scoping of data, which can refer to individual pieces of data (e.g. single fields in a relational database), the structured files where data are combined, the metadata, the information contained in the data, the software processing them, the algorithms underlying such software and the resulting knowledge/insights derived from the data.

The same **concept of "ownership"** is not clearly defined: there is not an official legal definition of it. Therefore it is opportune to adopt a working definition of the same. The concept of ownership refers to a legal instrument to allocate goods or rights to one or more persons, allowing them to exclude other persons from taking certain actions in relation to such goods or rights. Data ownership is an alienable legal construct allowing one or more persons (the owners) to control access to or use of a single piece or set of data, excluding others. Data ownership is alienable, not absolute and unlimited, since legislation may have an impact on the ability to control access to or use of the data. This means, for example, that despite the ownership of digital data, one has to comply with data protection law, avoid unfair commercial practices and destroy data subject to retention obligations.

Moreover, there is **legal uncertainty surrounding data ownership** in the manufacturing domain in relation to data produced by machines or devices, as well as non-personal data (e.g. in the area of finance). The uncertainties about the **concept of "ownership"** of data might represent a barrier for the data economy in Europe. However, the barriers to access and (re-)use of data are perceived as far more important. The reasons rely on many factors, such as, first and foremost, companies' worry of sharing sensitive information and losing their competitive advantage without even realising it: several companies do not feel confident in sharing its data due to this risk. Companies might have multiple reasons to be wary of sharing data (beyond what is legally binding) with other downstream players in the same value chain, which could be competitors, for the question of uncertainty about "ownership", usage of the data and of what others will do with the data.

For what specifically concerns **machinery data in global value chains and industrial data-driven industrial production and related industrial platform**, which are relevant to AI REGIO, data might be generated by component suppliers, customers, and other sub-contracted service suppliers: this happens, for example, through sensors in the production chain, in the after sales services, or through additional services, with smart machines coordinating manufacturing processes by themselves, smart service robots cooperating with people on assembling the products, and smart (driverless) transport vehicles covering the logistics side on their own, within the entire life cycle of a product (from concept to development, manufacturing, use and maintenance - and on to recycling).

Data generated in the Industry 4.0/Industry 5.0 dynamic value networks can be analysed by a third party service provider and be sent back to the respective client; on the other hand, they can also, once anonymized, be sent to the whole sector platform community or other third parties (such as sector regulators or benchmark designers) within the global value chains, potentially characterized by production sites scattered around the world. In this landscape, **the free movement of the different types of data originated at each step of such global value chain is essential** to any efficient production process, also remotely monitoring and maintaining the machines including the transfer at least some rudimentary data across production sites and most likely across countries.

Looking at the **types of actors along the manufacturing data value chain**[13] [12] and their respective contribution to it, they include:

- **Component manufacturers**, for instance of individual sensors or actuators (such as powertrains or driving shafts): they can generate data which are usually fed into the machine in which these components are employed. Currently such actors rarely gain access to the data across various machines which use their components. Though in the future such access

[13] Martina Barbero, Diana Cocoru, Hans Graux and others, "Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability. Study prepared for the European Commission DG Communications Networks, Content & Technology by Deloitte", 2017

appears to be technically possible, it is unlikely that the component manufacturer will be widely interested to it, given the knowledge which could be inferred by such process data;

- **Machinery Manufacturers** (manufacturers of the machines): since the machines are commonly equipped with various sensors and actuators enabling precise control, monitoring and optimization of production lines and remote maintenance is useful, such manufacturers have wide access their machinery in a remote way and gain access to relevant data;
- **Developers of production lines need access to capabilities** (including data collection and specifications of the machineries to be deployed): in case they also offer to run and maintain production lines as a service, they will have access to relevant data needed for such task.

As regards the **types of data generated** [12]**[14]**, potentially provided by, exchanged between and used by such players in the value network, each business users can provide different types of data, with added value in relation to the optimization of the production process and of the actual use of the industrial machines.

Different types of data originate at each step in the typical global value chain, from small components or even individual sensors or actuators over individual machines up to complete assembly lines planned, configured, and put to work by specialized companies. Commonly, the own machines are mixed/matched with machines from third party vendors in order to cater the clients' specific requirements. Notably, the types of data are as follows:

- **Original Equipment Manufacturer (OEM)**: the manufacturer of the original industrial machine (original equipment) assembles and installs parts supplied by subcontractors and then sells the finished product, customizing designs based on demand: the data provided by the subcontractors are widely used for optimizing and defining service level agreements;
- **Subcontractors of OEMs**: the third-party suppliers of products, components and/or modules provide data about their processes and products, to be used by the OEM to steel their processes;
- **End-users**: these are the machine owners and they generate data throught their production activities on their premises (production data) through machine parts and additional sensors deployed in the factory;
- **Third-party providers of IT infrastructure or data analysis**: they provide solutions through which OEM, their subcontractors and after sales service suppliers can collect and analyze the data and processes through algorithms, which create the added value for the value chain stakeholders in order to use such insights for process optimization and product enhancement;
- **After sales service suppliers**: they can collect data generated by machines on their operations, maintenance and repair needs: this kind of data is important for both the machine owners and the manufacturers, which desire to improve the performance of their machines.

Focusing on **legal perspective and the legal basis of the ownership claims**, many questions arise[15] [12].

These **questions raised by the data economy about who "owns" data and what data "ownership" entails and which kind of protection should be sought are difficult to answer** and there is not consensus around it. In other words, their commercial value and economic importance has inevitably led to calls for an ownership right in data, whilst it is key to strike a balance between competing interest at stake, including also the rights of data "owners" and the public interest in access to and reuse of data. In this direction, it is important to remark that it is critical to identify the **powerful basis for control associated to the ownership rights**: this means, for instance, that the data owner can provide access to data, restrict access partially or entirely, impose conditions on access or use, including charging fees. However, if data ownership rights are recognized, this power

[14] Martina Barbero, Diana Cocoru, Hans Graux and other, "Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability. Study prepared for the European Commission DG Communications Networks, Content & Technology by Deloitte", 2017

[15] Teresa Scassa, "Data Ownership", CIGI Papers No. 187, 2018.

to control data must have limits as well, for instance due to the need of balancing public and private interests: such rights must include exceptions that are functional to access and use data for innovation, knowledge, transparency, accountability, expression and privacy. This happens, for instance, with the IP Law.

The main challenges pertain to the **definition of the data** for the purposes of their application, locating ownership especially considering that data are often something in which there can be multiple interests (even in the case of personal information), whilst ownership rights seem a blunt tool to address competing interests potentially arisen across all contexts. Some advance consideration of a number of complexities is necessary: for instance, how should the data ownership right reflect factors such as the interests of a company collecting personal information and the interests of the data subjects in their personal information collected by such a company? Should the right be based on the source of the information or on the investment of resources in defining the parameters of and harvesting that information? In a smart manufacturing environment, how properly balance the interests of a company supplying the hardware that captures data, the company that derives data/insights from the captured data, the source of any other data used in the process of deriving new data, and the company that provides access to its equipment, machines and spaces in order to collect the data? It would be necessary further consideration and elaboration of issues like these, including the users' rights and the need to accommodate the broad public interest in access to and use of data, as well as of the circumstance that there can be so **many competing interests in data**, not just in the collection or creation of the data but also in its use. Besides the ownership rights or interests associated with data, in some cases, there are different rights and interests. In practical terms, ownership rights are frequently asserted in data, despite the nature, scope and robustness of such rights might be uncertain and contingent. In several cases, claims of ownership are asserted under the laws of confidential information.

More broadly speaking, **often the data ownership relates to a very diverse set of claims**: different regulatory regimes are evoked, such as intellectual property rights, data protection, trade secrets, contractual restrictions and other legal claims, and it is not well-identified the surrounding legal basis. In case claims would cause a **court dispute**, the result may differ depending on the legal basis, but also on the dataset, the country, the court and the sector concerned.

Different approaches emerged regarding the possible legal basis for data ownership claims, as follows.

I. **Confidential information/Trade secrets**

Data may be protected as confidential in certain circumstances, described in article 39(2) of the TRIPS Agreement. Information can be protected as confidential information if it "(a) is secret in the sense that it is not, as a body, or in the precise configuration and assembly of components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question; (b) has commercial value because it is secret; and (c) has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret." Despite some authors argue that data that are kept confidential can be protected as confidential information, it is also highlighted that such data are vulnerable to exposure through hacking or leaking by third parties and that, in many instances, significant amount of data cannot be kept confidential. However, confidential information has a non-proprietary nature and its protection is not the equivalent of an ownership right: this is due to the fact the value of confidential information typically lies in its confidentiality, and not in the information itself and, once confidentiality is lost, the information is often valueless. The benefits of relying on the law of confidential information to protect data includes the potentially infinite duration of protection, the breadth of subject matter protected, and the relative paucity of public interest exceptions permitting access or reuse. However, the main disadvantages regard the fact that not all data can be protected as confidential information and that, once confidentiality is lost, the protection is effectively at an end.

II. Copyright in Data

The existing frameworks for IP ownership are often called upon to provide some kind of protection for rights in data: copyright law represents therefore an important source of contemporary claims to ownership rights in data. This position relies on copyright law, which provide protection for compilations of facts that meet the threshold for “originality”, which is a standard applied to works differently, depending on their nature. In case of copyright, the author of a work is automatically protected (without need for registration): the right is automatic and of considerable duration. This form of protection is available for literary, dramatic, artistic and musical works, which are all defined broadly and also can be “compilations” consisting of multiple works of the same or different kinds, as well as compilations of fact. However, despite the categories of works are interpreted broadly and inclusively, this broad concept is mitigated by the further requirement that a work must be original and by numerous exceptions, including fair dealing. There is no protection for ideas as such, but only for an original expression of ideas. The issue of whether some data are capable of protection under copyright law is still under debate. Many scholars point out that the copyright law framework is ultimately inadequate for the task of addressing data ownership in a big data economy, since in the industrial data, generated by vast networks of sensors that observe and record the smallest units of entire global supply chains, the creative artistic features that copyright law serves to protect are missed. The data are far from creative in nature. On the other hand, facts, which are the building blocks of knowledge and innovation, are not protectable under copyright law: the copyright protection applies only to their original selection or arrangement and does not extend to the underlying facts. In fact, one of the principles of copyright law is that facts, which are an essential component of expression, are in the public domain largely because there is a strong and complex public interest in such facts being free to fuel new innovation or knowledge creation, in addition to the individual interest in being free to exchange and share facts without risk of legal constraint. The concept of data might be more complex than facts, and might involve more human agency: however, also data still fuel innovation, creativity, research and expression and therefore data monopolies would risk running counter to the public interest and might stifle both innovation and expression. In the case of compilations, the authorial contribution of the person who created the compilation that is assessed is what copyright seeks to protect: the originality requirement lies in the work of its author in selecting and arranging the different elements it contains and the threshold for originality is not particularly high. Database rights offer a robust protection for compilations of data, though they do not cover the facts that make up such compilations. In Europe, also database protection laws might play a role in relation to data ownership claims, since it was created a *sui generis* database right through the European Database Directive, which seems to protect the investment made in structuring the data — in other words, in creating the database itself, rather than protecting a property right in data. However, there is a limited protection available for compilations of fact. Today the human and technological processes that underlie the big data economy overwhelm the focus on the originality of selection or arrangement: the data are collected ubiquitously and continuously, as well as processed, analyzed and stored in increasingly complex ways, with creation of new data in the form of profiles, predictions and analytics. Within this evolving context, it is still under discussion whether data are copyrightable and, therefore, if data are protected. Part of the doctrine underlines that there can be no copyright in either facts or ideas, but only in their original expression: in case the expression of a fact or an idea merges with that fact or idea, there can be copyright because it would give rise to a monopoly over the fact or idea.

The international treaty context could be relevant in this discussion. Article 2(8) of the Berne Convention suggests that while certain types of facts are in the public domain, others may not be: in fact it provides that “[t]he protection of this Convention shall not apply to news of the day or to miscellaneous facts having the character of mere items of press information.” Article 10(2) of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement), provides some arguments to the need of a more complex distinction between "mere" facts and the more complex concept of data, providing that: "Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such. Such protection, which shall not extend to the data or material itself, shall be without prejudice to any copyright subsisting in the data or material itself". This might also be interpreted as acknowledging the potential for separate copyright to exist in the elements of a compilation. Furthermore, the reference to "any copyright subsisting in the data" leaves open the possibility that it is possible that a separate copyright might subsist in the data, despite copyright in a *compilation* of data does not extend to the underlying data.

In copyright jurisprudence some decisions leave open the possibility that some data might be entitled to copyright protection and the case law is slowly evolving to adopt a more complex approach to rights in data. It does not categorically exclude the possibility that some data constitute protected works (being original enough) and, even if data themselves are not works, compilations of highly original data might be well protected, as they may demonstrate a strongly original selection.

Academia is still debating the relevance of intellectual property protection in relation to the "ownership" claims and to the access and (re-)use of data: the scholars have not agreed yet on the opportunity of using IPR regime to further stimulate B2B data exchange. However, companies do not seem to rely on IPR in this area and this is not representing an obstacle for sharing, accessing and re-using data, on the other hand.

However, while copyright might represent a basis for the protection of rights in data, it has also some drawbacks.

First of all, the copyright law evolves relatively slowly, and, even if the case law shows the capacity for the protection of data, it is unpredictable of copyright law will more in this direction. It was also proposed to reach consensus at the international level on whether, and to what extent, data should be protected under copyright law.

Furthermore, copyright protection requires a human author: in case of works created by machines/ automated processes, there is the lack of human authorship and, therefore, no copyright protection. These debates have relevance in the context of AI and AI REGIO — and by extension to contexts where data are generated through AI: they regard the output of artificial intelligence (AI) processes, which should not be capable of copyright protection.

III. Sui generis right

Other authors raised the possibility of creating sui generis ownership rights in data. The debate over the need for enacting this broader sui generis data ownership right, which is mainly concentrated in the EU, seems comparable to the 1990s discussion on the need for sui generis database protection, which brought to pursuing this route in Europe (not in the US and in Canada). This means to create a new data ownership right. However, the challenges are somewhat daunting and include, for instance, the definition of the data in which ownership rights can subsist, as well as the definition of appropriate rules for allocating ownership (which adequately take into account the complex ways in which data may be co-created) and maintaining an appropriate balance between, on the one hand, the ownership rights and, on the other hand, the public interest in fostering innovation, and supporting research, criticism, free expression, education and creativity.

The establishment of a new regime in this rapidly evolving data environment might reduce flexibility and increase complexity. There are the risks of unduly burdening an industry that has thrived on fast-paced and flexible innovation and to neglect users' rights, without granting fair rights of access and reuse of data and carefully assessing the needs of owners and users of data.

Some doctrine argues that it might be preferable to allow existing law to continue to evolve, due to the challenges and complexities around data, ranging from their definition, to the locating of their ownership and balance competing interests at stake, as well as to the need to set significant rights of access and use: this involves the evolution through the application of established principles, IP regimes and, above all, commercial agreements.

IV. Contract law

Practically all data are gathered purposefully and systematically: therefore individual contracts already cover data ownership, exchange, access to and use of data among the actors along the value chain. A large majority of experts commonly agreed that, in the lack of suitable legal instruments, these individual contracts are sufficient to manage the data ownership and the control of access and (re-) use of data. It believes that at this early stage an overarching legal framework for data ownerhip claims and rights would hamper innovation and slow down the evolving business models, rather harming than benefitting users of Industry 4.0 opportunities and industrial platforms.

Usually and in almost every type of industry, the data go to the manufacturer company, which usually is also the provider of the industrial platform cloud service has the digital control over the data (which is internal to the machine, a part of the machine itself).

In IoT and Industry 4.0 (and likely Industry 5.0) contexts, manufacturers often seem to want to control data within the boundaries of their machines but also beyond, i.e. via a platform. This can be in contrast with end-users' interest, which might fear that they have to pay extra for the access to data (even their own machine´s data) or give up control over their machines. In everyday business today the general rules and common practices existing in relation to the industrial platforms/cloud environment foresee that the files stored in a cloud are owned by the company who created the file and set it up, which usually is the "data generator". Nevertheless, this does not imply that the content of the file (text, data, etc.) is protected, for instance by copyright.

The data generator can give access to "machine data" on a contractual basis to anyone. On the other hand, someone argued that "machine data" can also be seen as "raw data" which do not belong to anyone and should be treated like a resource: in this case, OEMs and cloud service providers could be willing to lead to a proactive storage of machine data in order to keep all options of re-use open for the future, whilst, SMEs (suppliers of parts etc.) do not share this view and see their trade secrets and confidential information infringed and claim to have the right to access and use "machine data" as well.

There are still many challenges and concerns around data exchange, data ownership, data innovations and customer demands, also regarding security, sensitive data, and technical reliability.

However, as mentioned, the contractual approach seems to currently be the more adopted and it is often link with technical solutions capable of enforcing it.

In this direction, it is important to mention the **data sovereignty** concept and standard[16] [14] [15].

It is recognized by the practitioners that the data sovereignty concept offers the opportunity of sharing data in a secure and sovereign manner supporting the manufacturing industry in several processes or segments, like self-registry and completing a digital twin with manufacturer information and data centered business models and services. Data sovereignty provides the trust and security between partners in a data centered ecosystem and the data sovereignty between them. Data sovereignty relates to both access control and usage control. The owner can decide with whom, how long and under which conditions he wants to share his data and he also can control the further usage of the data, once the data has been accessed. Data sovereignty is considered by many the key link between, on the one hand, the creation of data (in the IoT environment) and, on the other hand, the use of such data in machine learning (ML) and AI algorithms. It implies the ability to describe, trade and protect the asset data. It is capable of providing an answer to market inhibiting effects of data economics and in

[16] IDSA, "Data Sovereignty – Critical Success Factor for the Manufacturing Industry", 2021; IDSA, «White Paper "Sharing data while keeping data ownership. The potential of IDS for the data economy,» 2018.

particular as regards the Industrial IoT, the AI and any kind of smart service scenarios. It is expected to be the basis for successful AI by making considerably more data sources accessible. Data sovereignty presupposes metadata attached to data, unambiguously defining data usage policies at each level of the data value chain and providing information to the receiver on the origin of the data and its circumstances. The technical infrastructure should be able to enforce data sovereignty, facilitating through flexible and pragmatic solutions the execution of contractual provisions on the use of data. These provisions can enforce the data policies in terms of processing, allow (or disallow) linkage or analysis of data-by-data users, or allow (or disallow) third parties access to data, and other use limitations, flow control, data transfer restrictions, etc. In case of need in AI REGIO environment, data sovereignty should be ensured also within its wide ecosystem, including third parties' digital infrastructures (e.g. clouds, software components, networks).

The successful data sharing in industrial contexts could therefore rely on the data sovereignty concept as described by the International Data Spaces (IDS): this would enable the manufacturing companies to retain control over the collection and usage of their data and, as a consequence, to scale and grow with data. The benefits include aspects like to keep the competitive advantages, to efficiently fulfill customers' expectations and to define and implement innovative business models and services with the trusted use of more data. This governed and decentralized approach is promoted to make the most out of data and to give experts access to them in a traceable, evolutionary environment in order to gain insights about data and their interdependencies.

In this environment, IDS-RAM and DIN SPEC 27070 standard[17] [15] should be monitored: the former is the reference architectural model for data sovereignty, used when interchanges are desired to be carried out maintaining the property and governance of those items to be exchanged (data, models, etc.). The latter is the IDS standard, published on February 21st, 2020: "Requirements and reference architecture of a security gateway for the exchange of industry data and services". In such a standard, parts of the current version of the IDS reference architecture (version 3.0) has been incorporated for operating secure and trustworthy infrastructures for data exchange. It should be considered in the perspective of guaranteeing data sovereignty, in order to incentivize the data sharing and build/reinforce trust among participants.

In relation to the contractual context, it is important to recall also the **Industry Agreements (IAs)**. The European Commission, Directorate-General of Communications Networks, Content & Technology commissioned a study on them as a tool for stimulating the activities in terms of creation and development of European data spaces. The Industry Agreements (IAs) "are bi- or multi-lateral (voluntary) contractual frameworks/model agreements, designed to support the development and functioning of such data spaces. Notably, they are designed to address the different building blocks required to develop an industrial data space – technical specifications (e.g. IAA, exchange protocols), data specifications (e.g. data structure, semantics) and governance and legal dimensions – encompassing all possible stages of the data life cycle. Such agreements can be regarded as a strategic tool to guide industry players' efforts by providing a taxonomy of key aspects to look into for the adoption of common rules by industry actors, the creation of voluntary B2B data sharing schemes; and innovation in general. Indeed, industry agreements provide the high-level classifications of the core elements and contractual clauses that are needed to reach a sufficient level of data interoperability, exchangeability, and quality"[18] [17].

The "industry agreements" (IAs) are seen by the EC as key to set the European industries' common understanding on functionalities, architectures, specifications, interfaces for the

[17] IDSA, «DIN SPEC 27070,» 2021.

[18] European Commission, Directorate-General of Communications Networks, Content & Technology, "BACKGROUND DOCUMENT – Workshop on recommendations and impact validation - Technological and economic analysis of industry agreements in current and future digital value chains. A Study prepared for the European Commission. DG for Communication Networks, Content & Technology by CARSA, Ecorys, KU Leuven and VDI VDE IT", 2021

platforms, reference architectures and interaction protocols capable of supporting the growth of ecosystems and standard development. Among the more significant areas of innovation, which could generate a high economic impact for the data-driven European economy, there is the data sharing/exchange. During the final webinar conducted in the context of this European Commission's (DG CNECT) study, the barriers and opportunities for such Industry Agreements were outlined, with a focus on how they can be initiated and how they can have a role in the development of the data spaces ecosystem and in the advancement of the industrial digital ecosystems for the benefits of European businesses by reaping the benefits associated with digital innovation[19] [18]. Further details in relation to the Industrial Agreements, the outcomes of the Study and their potential relevance for AI REGIO will be investigated in the next release of the document, D7.2 "AI REGIO Human-AI Interaction Framework".

Currently, there is strong reliance on contractual tools for sharing and accessing data and it is likely that these tools will remain the key vehicle for organizing and regulating commitments within the Data Economy[20] [12] [12]. This implies that data ownership, as well as data access and (re-)use, will remain defined with **pragmatic and de facto arrangements on a case-by-case basis** and through bilateral relations. This might pose **some challenges for SMEs**, since they might not necessarily be able to bear the costs of such an approach and they might lack the commercial, negotiation and bargaining power to get access to the data they need. SMEs would face disadvantages which would hamper innovation by these players. It would be advisable the adoption of a legal instrument homogeneously defining the data ownership and respective usage rights for up- and downstream businesses, as well as the relationship between businesses and users of a product or service: it would provide a level playing field for businesses (SMEs and large enterprises) to develop their business model and find their niche in the data economy.

Furthermore, the **implications for personal information** should be considered when discussing on data ownership, in particular introducing adequate consent management tools and models. GDPR already recognizes the individual interests in their personal information, despite they currently stop short of ownership rights and the evolving frameworks are still moving outside a formal ownership regime. In case of enlistment of new ownership rights in data but also in case of adoption of the contractual approach also in the future, it will be difficult to exclude individuals from ownership of the personal information they generate simply by using the machines or acting in the plants. AI REGIO Consortum will take this aspect into account, both in the development of its technology and in its experiments, as also envisioned in the legal and ethical requirements set in D2.7.

3.3 Privacy, Data protection and other ethical issues

Other legal and ethical challenges relevant to AI REGIO, and especially to the human-machine-interaction technologies and human-centric aspects of AI-based manufacturing systems, arise.

They range from Data Protection and Privacy, to the risk of stigmatization and social sorting, to the need of respect of human rights. Other ethical dilemmas include, for instance, the need to avoid or at least minimize algorithmic biases, as well as to ensure that safety and security concerns are properly addressed are critical. Employees' comfort and well-being must be prioritized, including adequate consideration of the psychological issues potentially brought by the CI-driven working

[19] European Commission, Directorate-General of Communications Networks, Content & Technology, "Introduction to the Final webinar & DRAFT Executive Summary - Technological and economic analysis of industry agreements in current and future digital value chains. A Study prepared for the European Commission. DG for Communication Networks, Content & Technology by CARSA, Ecorys, KU Leuven and VDI VDE IT, 2021.

[20] Martina Barbero, Diana Cocoru, Hans Graux and other, "Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability. Study prepared for the European Commission DG Communications Networks, Content & Technology by Deloitte", 2017; Teresa Scassa, "Data Ownership", CIGI Papers No. 187, 2018

environment, such as the risk of "technostress" and, especially, the risk of development of emotional attachments to machines. It should be investigated whether these emotional or social bonds between humans and machine imply a subtle form of deception related to the subconscious processes involved in human–robot interactions and, in case yes, whether this is ethically acceptable and could contribute to the good life/well-being.

These aspects will be deepened in Sect. 4.2 "ELS dimension" of the CI paradigm.

4 Human Factor within AI REGIO Collaborative Intelligence Model

4.1 Human-centric CI Approach in AI REGIO: "humans in the loop" train-explain-sustain paradigm fostering data-and-human-oriented SME digital transformation

As stated in D5.1[21] [18], the new paradigm Industry 5.0 is appearing on the horizon, although the Industry 4.0's push has not yet been exhausted. In particular, this new paradigm aims to take into account manufacturing machines with workers in the loop with the final aim to realize a synergistic collaboration between humans and machines. To achieve this collaboration, it is essential to look for the right balance between high value-added tasks (requiring human brainpower and creativity of humans) and repetitive tasks (requiring high speed, precision, and security typical of the machines). Through such a balance, humans and machines can enhance each other's complementary strengths. Under these conditions, humans must provide creativity, teamwork, and social skills, while machines can augment human capabilities with speed, precision, and scalability. Thanks to this evolution, the humans will work side by side with machines, not only with no fear (overcoming the Frankenstein complex) but also with serenity, knowing that their co-workers (i.e. machines) have now become trusted collaborators.

The above-mentioned fruitful humans-machines collaboration can be achieved only if the machines will have a full understanding of human behavior, including intentions, emotions, and desiderata (cognitive ability). In addition, the use of a machine in real environments will require the control of the operation field and of the surrounding environment (perception ability). In these contexts, machines will also need decisional autonomy, which is a key feature to make them useful in real applications and to allow them to fruitfully collaborate with people also in dangerous and unstructured environments. In the field of robotics, this interaction between humans and robots (interaction ability) will constitute an important aspect for current robotic applications aimed at assisting people, instead of replacing them. In the definition of these new human-machine interaction paradigms, safety and ergonomics will be the major issues that must be considered.

An interesting idea to implement I5.0 within the AI-REGIO project comes from the concept of Collaborative Intelligence (CI) reported in the *Harvard Business Review "*Collaborative Intelligence: Humans and AI Are Joining Forces" [19][22]. The authors state that "Most activities at the human-machine interface require people to do new and different things (such as train a chatbot) and to do things differently (use that chatbot to provide better customer service). Organizations that use machines merely to displace workers through automation will miss the full potential of AI. Tomorrow's leaders will instead be those that embrace collaborative intelligence, transforming their operations, their markets, their industries, and—no less important—their workforces".

[21] Gianfranco Modoni, Marco Sacco, AI REGIO D5.1 Collaborative Intelligence and Industry 5.0, 2021

[22] Harvard Business Review https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces

In this report, Collaborative intelligence envisages two bidirectional collaborative interaction patterns, which are explained in the two following sections.

4.1.1 Human-to-machine: train explain sustain

The first direction "Humans Assist Machines" goes into the direction to support the ways through which humans can augment machines. The idea is that humans can provide this support in three different ways: humans train machines to perform specific tasks, humans explain the results of those tasks, and finally, humans sustain the responsible use of machines. This support can be brought back to the metaphor of a Parent-Child family relationship. In this relation, indeed, "the parents aim at transferring as much of their knowledge to children (train), at understanding their viewpoint and establishing a constant positive dialogue with them (explain) and in taking a collaborative and not punitive attitude when they make mistakes (sustain)". In addition, just like family's fathers, humans should consider machines“ as their own creatures” by establishing a relationship with them. The latter can bring a further benefit since the more humans perceive machines as their digital heirs, the more AI will be accepted in next-generation workplaces.

4.1.2 Machine-to-human: amplify-interact-embody

The second pattern "Machines Assist Humans" aims to amplify and extend both cognitive strengths of the humans and their physical capabilities, and to act as a digital twin of the person which allows to interact with other humans (e.g., customers and employees). And the interaction declined by this pattern is similar to the Caregiver-Elderly relationship, where the caregiver aims to raise the elderly cognitive and physical capabilities when the latter are insufficient.

To implement this pattern, it will be essential to understand how machines can support human capabilities and heighten their creativity (for example AI can support the human decision making process by providing the right information at the right time) and how to redesign business processes to enable the humans machine collaboration. A fruitful collaboration between humans and machines will also require a valid framework that defines the rules of this interaction. Indeed, modern companies are looking for support and guidance on how to overcome the ethical issues that are associated with AI.

Humans therefore should see Machines as their capability multipliers, making it possible unprecedented and innovative activities at cognitive, physical and relational level.

Since we have seen the CI is envisioned in two directions, which direction can be of interest and mut be implemented in the scope of AI-REGIO? The answer is, in both the directions. The direction to be taken is largely depends on the case studies

4.2 The bidimensional nature of HF in AI REGIO CI solutions

The AI REGIO project is moving ahead for facing the challenges posed by the paradigm shift towards Industry 5.0, with AI-driven autonomous systems efficiently and effectively interacting with Humans according to the Collaborative Intelligence paradigm, thereby dealing with the human-centric aspects of AI-based manufacturing systems and enabling an immersive AI-based digital workplace. The interactions with information, processes, machines, and people will be made simpler also via Personal digital assistants enabling AI REGIO technologies to interact with an organization's digital workplace to achieve desired outcomes.

Human factors and ergonomics, commonly referred to as human factors (HF), is the application of psychological and physiological principles to the engineering and design of products, processes, and systems.

These breakthroughs entail, on the one hand, a technical dimension and, on the other hand, an ethical, legal and societal (ELS) one. Both of them will be described in the following paragraphs.

4.2.1 Technical dimension

A new information platform should drive human behavior in the Industry 5.0 oriented workspace of the future. This platform should exploit the full potential of the collaborative intelligence between the human and the manufacturing system, according to the interaction channels from sections 4.1.1 and 4.1.2. This platform should not be a simple digital assistant behind the workers, but it should mainly support a more efficient, safer, even proactive interaction with the workspace. Under these conditions, it "should act as the worker's digital counterpart, looking inside-out to reflect all the aspects of the worker, including the interaction with the environment, skills, preferences, and even the mood, fear/excitement".

Within AI REGIO, a platform that meets the above requirements has been introduced in D5.1. In particular, as Industry 5.0 aims at merging computational intelligence (and cognitive computing capabilities) with human intelligence in joint operations, this platform provides the functionalities to refine these collaborative interactions between humans and machines, leveraging a real-time monitoring solution of the human-centered processes. This monitoring will enable checking the efficiency of each designed process, updating decisions about resources management, predicting failures, and assisting in implementing optimization strategies. This way, the platform for real-time monitoring promotes harmonization and orchestration between machines and human factors, especially considering the cognitive and physical workload. In addition, the platform enables the comparison of these different solutions according to scenarios' needs.

Since Digital Twin (DT) is becoming a consolidated technology to simulate the physical industrial asset performance, allowing to predict failures or investigate problems, the AI REGIO platform leverages the potential of a DT-based solution for optimizing human-centered processes. The DT represents a virtual and faithful mirror of the physical process that allows monitoring the process parameters, the comparison with analytic models, and the feedback, in real-time, by setting parameters to keep the process in optimal conditions. However, it should be noted that the traditional model of DT, which does not consider the worker's presence, cannot be compliant with the concept of Industry 5.0, which instead foresees the centrality of the worker. To bridge this gap, the AI REGIO platform proposes a new DT conceptual model compliant with industry 5.0, to be adopted as a reference architecture for manufacturing companies. In addition to the framework, task 5.1 of AI REGIO project provides to implement a new DT adhering to the proposed conceptual model's specifications. The new reference architecture will also be validated within AI-REGIO project in a real case study, to demonstrate the correctness and benefits of the overall proposed approach.

In this regard, in order to implement this platform, various technologies are required, divided by their roles in supporting human-machine collaboration:

- (Machine assisting humans), enabling human work (physical, mental) through an efficient decision-making and execution support. The examples are various types of decision support systems, knowledge-based and cognitive systems. One major challenge to implementing this interaction is adapting the (physical and virtual) machines to the needs of each person; it should account for human diversity.
- (Human assisting machines), enabling an efficient learning process (by machines), which will provide models for automated decision-making. An example is various types of supervised learning methods, where humans prepare datasets for machine learning. Another form of Humans-Assist-Machines is when humans provide Expert Domain Knowledge under the form of models.

- (Human assisting machines assisting humans) collaboration continuum, where the humans support the creation of models (by machines), used in the decision-making support (for humans). These decisions can create data used in refining the models.
- Novel learning paradigms that accelerate knowledge acquisition for machines can boost the above Industry 5.0 scenario, including active learning and transfer learning, as well as their combination with unsupervised learning. Moreover, explainable AI technologies (including explainable robots) will boost the transparency of robotic operations and their acceptance by humans. This will contribute to the overall trustworthiness of the human-robot collaboration

4.2.2 ELS dimension

The AI REGIO project is moving ahead for facing the challenges posed by the paradigm shift towards Industry 5.0, with AI-driven autonomous systems efficiently and effectively interacting with Humans according to the CI Model, rotating around the human-centric aspects of AI-based manufacturing systems and enabling an immersive AI-based digital workplace. As highlighted in section 4.2.1, the interactions with information, processes, machines, and people will be made simpler also via Personal digital assistants enabling project's technologies to interact with an organization's digital workplace to achieve the desired outcomes. In AI REGIO, Industry 5.0 AI-empowered technological artefacts under development are converging in expectations of collaborative services and applications enabling more widespread use of CI solutions within a human-machine co-working environment.

Considering the **pivotal nature of work in most adult lives**, the underlying ethical, legal, regulatory, psychological and societal impacts of Industry 5.0 and CI solutions might be more profound in respect to other sectors, such as transport. The adequate consideration of these impact is paramount in our trajectories towards the Industry 5.0 CI-driven paradigm, in order to ensure, on the one hand, that both industrial companies and workers benefit from the digital transition exploiting the advantages of a synergistic collaboration between humans and machines, reciprocally enhancing each other's complementary strengths and, on the other hand, that the workers (and their rights) are put at the center of the factory.

Therefore, among the three core elements of Industry 5.0, as depicted in D5.1 "Collaborative Intelligence and Industry 5.0"(human-centricity, sustainability, and resilience), from the ELS it is necessary to especially **focus on human-centricity**. This puts the cutting-edge technologic advances at the service of human needs and interests in view of adapting the production process to the needs of the worker, going beyond the improvements in the workplace and in the production processes and involving the workers (together with the customers) in the end-to-end loop of the production process[23] [18]. It is acknowledged by AI REGIO Consortium that **human well-being has to be placed at the heart of the development and operation of its technologies and solutions**: the human-centered technologies are rapidly spreading in manufacturing and industry, due to their intrinsic potential at assisting workers and improving their working conditions, as well as to the new possibilities for humans to interact with smart systems and delegate some of their tasks to these them, with human cyber-physical system supporting the extension of human operator physical, sensorial, and cognitive capacities and capabilities.

As regards the main enabling technologies related with Industry 5.0, in this document the attention is especially directed to AI and **human-machine-interaction technologies**, interconnecting and combining the strengths of humans and machines, including, for instance, multi-lingual speech approaches, human intention prediction, cobots (collaborative robots) or augmented reality tools, as well as on wearables and exoskeleton and other CI tools characterizing distributed, less controlled

[23] Gianfranco Modoni, Marco Sacco, AI REGIO D5.1 "Collaborative Intelligence and Industry 5.0", 2021

work-settings. They raise several concerns in terms of safety, security and certification of the devices. Furthermore, the collection, processing, manipulation, transmission a of data, and eventually the decisions computed by AI algorithms pose likewise additional risks, for instance as regards privacy, data protection, security, liability and other.

Aspects like the **different personal preferences** toward working with machines and robots, types of robots/machine preferred to work with, learning to work with them, competing or cooperating with the same, and possible negative attitudes should be adequately investigated. It is essential, besides conducting a survey on the expectancy and vision, to perform experimentations on this aspect in view of moving ahead towards the **ethically-sound and human-centered definitive human-machine co-working environments operation** in the near future. The **AI REGIO TERESAs** are expected to provide a high-level contribution in this direction.

In the workplaces of the future, including those which will be experimented in a limited scale and in a safe and controlled environment in the next stage of the project through TERESA, humans and machines will share physical spaces working not only sequentially but even with close, physical with real-time responses from machines/robots to users, thanks to more complex sensing and control applications. These might, for instance:

i) **Human supervisory control of machine/robots in performance of routine tasks**, with handling of parts on manufacturing assembly lines and accessing and delivery of packages, components, other objects in the plant. In this scenario the machines are capable of performing a limited series of actions automatically, based on a computer program, besides sensing its environment and its own joint positions and communicating such information back to a human operator who can update its computer instructions as required;

ii) **Remote control of space and vehicles for non-routine tasks** in hazardous or inaccessible environments, with manipulation and mobility tasks in the remote physical environment in correspondence to continuous control movements by the remote human. The machine is intermittently reprogrammed by the human supervisor to execute pieces of the overall task;

iii) **Human–robot enahnced interaction**, including robot devices to provide teaching, comfort, recommendations and assistance for the human operator.

In order to reach the best ethically-sound human-machine manufacturing collaborative system, it is essential to provide humans and machines with adequate abilities to cooperate with each other, the so-called **Know-How-to-Cooperate (KHC)**[24] [21], which can be considered as a way of maintaining humans in the loop and fully benefiting from technological abilities by keeping humans in the supervision and control loop and by maintaining their situation awareness and ensuring that the human stays involved in the decision-making process.

Besides questions on the acceptance of robots in the workplace, discrimination against robots or people, achieving privacy and trust in a human-robot collaborative work environment, additional **subtle ethical challenges** might be brought beyond the immediately obvious, such as the need in a human-machine interaction to balance the competing interests at stake of many different actors (users, managers, designers, customers, etc.). Other key questions regarding the impact of this technology on humans and on society need to be further investigated: though some of them have already been addressed (though only narrowly), broader questions regarding ethical issues, especially focused on the human employee's perspective, personal implications and societal impact, have rarely been in-depth considered in this field. Human-Machine-Cooperation principles might serve as a guide in this direction to handle the ethical risks and to find the best balance between the competencies and roles of humans and machines within the environment in which they have to make decisions, potentially with other decision-makers involved in this workplace, integrating machines and robots.

[24] Marie-Pierre Pacaux-Lemoine, "HUMAN-MACHINE COOPERATION: Adaptability of shared functions between Humans and Machines - Design and evaluation aspects", 2020.

In the following paragraphs, some of the **main ethical, legal, regulatory, psychological and societal impacts of Industry 5.0 and CI solutions** will be investigated in order to move forward to achieve **Trustworthiness.** These impacts complement the topics and remarks indicated under Sect. 3.

4.2.2.1 Data Protection and Privacy

In a Human-machine CI workplace it is paramount to ensure i) that participants are aware that their data are being collected, ii) that they give informed consent to this, iii) that data are stored in a secure manner, iv) and that the ethical requirements for data collection, storage and use are met.

Depending on the environment, individuals expect different levels of privacy and such level decrees in the public space: nevertheless, in the workplace, which is considered as a type of public place, **employees still expect a certain level of privacy**. However, the concepts of privacy and trust could require some evolution as CI tools and robots become part of our employees' daily life and workplace: it should be taken into account also that industrial devices and machines are capable of recording everything, being often equipped with sensors and they will likely be able in the future to read minds using electroencephalogram (EEG). For instance, cobots are able to collect data to adapt to the abilities, work-rate and needs of their human coworkers. In AI REGIO and in line with the concept of Industry 5.0, the full exploitation of the potential of the CI paradigm along the human-manufacturing system interaction entails the data collection, processing and use in relation to the real-time monitoring solution of the human-centered processes, as well as to the worker's digital replica/twin to capture worker's skills, preferences, even the mood, fear/excitement (see above sect. 4.2.2) and the analysis of the past behaviour. This will allow to go beyond the commitments towards and improved interaction in the workspace and thereby to enable a continuous and autonomous improvement of the collaboration. The in-depth consideration of the human component within the **Digital Twin (DT) loop** or, in other words, the presence and centrality of the worker and the consequent Personal Digital Twin enabling to tailor the interaction modalities to the status, preferences and behaviour of a person, also requires for **additional personal data collection and use**. Among the data models that will be used in AI REGIO, the **Human Data Models** will support a coherent representation of Human-AI interaction, aimed at ensuring the successful human-centric engineering and adaptive automation that fits the specific needs of different employees (e.g. for novice, older and disabled people)[25] [18]. These Human Data Models will include, for instance "the human role, goals and tasks; demographics, key anthropometrics, functional (sensorial, physical and cognitive) capabilities; knowledge and skills; needs and preferences; physical, cognitive and emotional status (e.g., based on physiological measures) & dynamic behaviours"[26] [18].

Such data collected can be processed in cloud services, as well as both **the data themselves and the models/insights derived from them might be stored and used by other machines** within robot systems consisting in a network of distributed processes.

In this context where machines are able to capture data about people, besides about equipment and environments, the main challenge lies in using data ethically and in an informed manner. Nevertheless, **it might result complex to gather informed consent according to the European Data Protection Framework (GDPR)** and the underlying ethical principle for these data collection, processing and storage which can be seen as an operator monitoring. In any case, it is essential that the participants are clearly aware that data were being collected, handled and stored and the AI REGIO Consortium is committed in ensuring this.

[25] Gianfranco Modoni, Marco Sacco, AI REGIO D5.1 "Collaborative Intelligence and Industry 5.0", 2021

[26] Gianfranco Modoni, Marco Sacco, AI REGIO D5.1 "Collaborative Intelligence and Industry 5.0", 2021

4.2.2.2 Stigmatization and social sorting

It is acknowledged that **people can change their behaviour in the field when they are aware to be monitored or observed**. The AI systems, which often evaluates performance and are designed to better fit to the human co-worker, can be perceived as a treat[27] [22]: the workers could worry about being stigmatized for their performance. This implies the mental stress of being held to the productivity standards of a robot. Furthermore, there might be the tendency for work environments to be perceived as coercive. The workers' livelihoods might be related to the acceptance and accepting and adapting to technological change and the workers may feel under pressure/forced to conform to what the management asks, suppressing their desire to avoid surveillance/observation for the fear of being negatively evaluated, as if there is something to hide. In other words, the final end-users, the worker, may be implicitly pushed by the working environment (or other form of social pressure) to adopt the innovating technologies, thereby **not** giving rise to a **really voluntary use** of them.

The mental health risk for users that could be generated by the future "behaviour" of AI applications, for example in case of their collaboration with AI robots and systems in the working environments as foreseen in AI REGIO, is also mentioned by the EC's "Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics"[28] [3], which highlights that is should be covered within the concept of product safety in the legislative framework.

4.2.2.3 Psychological Issues

This consideration allows to move to another important topic, the phycological issues resulting from the Human-Machine interaction in the CI-driven working environment. It is widely acknowledged that cutting-edge technologies might have certain psychological effects on humans. Relying on this assumption, it should be considered that HM interaction within CI systems in industrial settings might have **unprecedented psychological effects** on the workforce.

A first psychological issues related to working with robots could be a sort of **"robophobia"** or a **"technostress"**, though it is expected to be not so likely in case of AI REGIO, where technological artefacts are going to be developed with a human-centric approach and for empowering humans.

Nevertheless, there could be the risk that workers experience increasing technostress, which might generate a decrease of their engagement and pleasure in performing their tasks, as well as of the satisfaction at work, feeling less useful and in control at work and not comfortable when working with a robotic partner. There might also be a sort of mistrust towards robots when the workers must share decisions with machines, since they might be perceived as antagonists, even in collaborative scenarios.

These elements have been considered psychosocial factors of stress increasing the risk of developing **musculoskeletal disorders** (MSD) at work. Though these disorders are often evaluated in relation to the exposure to biomechanical factors capable of generating them (high efforts and awkward postures), it is important to avoid increasing the mentioned psychological factors as well, leaving job control (such as the work pace) to the worker and preserving some "human added value" in one's job. It is related to the **mental health and well-being of the workers**, which relies also on **self-esteem**. This is exactly what AI REGIO CI system is going to do.

[27] Gordon Briggs, Matthias Scheutz, "How Robots Can Affect Human Behavior: Investigating the Effects of Robotic Displays of Protest and Distress", 2019

[28] European Commission, «COM(2020) 64 final. "Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics",» 2020.

Another aspect pertains to the **positions and roles robots and other CI machines will assume in the organization**.In Human-robot co-working environments where robots are placed in positions complementing what humans do, humans will likely not feel threatened by robots, because the machines will help humans in completing tasks without the burden of managing a human assistant, leaving more time for tasks requiring creativity and higher intelligence. On the contrary, it is not clear the reaction of humans in case robots may take positions that humans compete for. But this question arises especially for future scenarios, not for the deployment and use of AI REGIO technologies and project's experiments.

Two other ethical risks when designing and using CI human-machine systems in industry are the so-called **"Master-Slave dependency"** and the **Emotional dependency**[29] [21][30] [23]. Both of them need to be considered and properly tackled especially in the post-project phase, when AI REGIO outcomes will operate in a real working environment.

The former refers to the human temptation to delegate more and more functions and tasks to machines even without a real need but for convenience, curiosity or without paying attention to the consequences. The risk is that, doing so, the operators become progressively dependent on the machines and less and less train themselves to carry out the allocated functions and tasks, with subsequent loss of skills, which might be important in case a problem occurs.

The latter refers to the fact that there might be the tendency, especially if a machine is perfectly tuned and adapted to support humans like in CI-devices in Industry 5.0, to develop a kind of empathy and thankful emotion towards the machine, with the risk of creation of emotional dependence, accompanied by overconfidence in the ability of the machine to solve problems and facilitate one's tasks, even in unexpected situations. The operators might generate the attitude to find it easier and more enjoyable to interact with a machine, especially in a CI environment, rather than with humans, diminishing their social abilities.

In many human–robot interaction studies[31] [23] it has been reported that, despite existing robots are not sentient and lack feelings, nevertheless people have a sort of social responses to their presence and they behave as if they are **social entities**, arguing that robots' autonomy and agency could give them a form of **robotic personhood**, creating a new class of legal entity[32] [23].

More in general, in a CI workplace, where people and machines interact on a regular basis on the basis of a human-centered model in an increasingly complex and humanlike flavour, and with an increase of people's expectations about machines' capacities, some workers could start to develop some kind of relationship with the machines. The question is whether these **emotional or social bonds between humans and machine** could contribute to the good life/well-being.

Some authors coined the term of "**socially interactive robot**"[33] [25] meaning "a robot that elicits social responses from their human users because they follow the rules of behavior expected by these human users". In a CI-empowered workspace the robots are expected to operate in spaces specifically designed for humans capable of communicating and interacting in a sociable way in

[29] Marie-Pierre Pacaux-Lemoine. "HUMAN-MACHINE COOPERATION: Adaptability of shared functions between Humans and Machines - Design and evaluation aspects", 2020

[30] Marie-Pierre Pacaux-Lemoine, Damien Trentesaux, "Ethical risks of human-machine symbiosis in Industry 4.0: insights from the human-machine cooperation approach", 2019.

[31] Marie-Pierre Pacaux-Lemoine, Damien Trentesaux, "Ethical risks of human-machine symbiosis in Industry 4.0: insights from the human-machine cooperation approach", 2019.

[32] Maurice P, Allienne L, Malaisé A, Ivaldi S. "Ethical and social considerations for the introduction of human-centered technologies at work". In 2018 IEEE workshop on advanced robotics and its social impacts (ARSO). ieeexplore.ieee.org. 2018

[33] Maartje M. A. de Graaf, "An Ethical Evaluation of Human–Robot Relationships", 2016

order to ease the communication with its users, evoking social interactions (or just reactions), according to the rules of human social interaction behaviors.

This facilitates the development of **emotional attachments** to robots.

In addition to the functional requirements of robots performing their monitoring and assistive tasks in the social environments, and in particular in the workplace, some authors argued that this requires **socially interactive components** to engage in social interactions and create relationships with their users in order to achieve their goals.

In conjunction with this, there is also the concern whether **this form of attachment to robots might replace human contact all together** contributing to **social isolation** and to diminished willingness to deal with the complexity of real human relationships**:** social interactions with other human beings give meaning to life and guidance to appropriate behavior, besides being essential for the development of the social self.

Furthermore, especially in case of human–robot interactions constructed along humanlike interpersonal interactions, it should be investigated the human attitude to form unidirectional emotional bonds with these technological artefacts and if it is ethically acceptable this **subtle form of deception** related to the subconscious processes involved in human–robot interactions and it should be discussed whether potential bonds between humans and robots could contribute to the operators' well-being.

These aspects advocate that work ethics evolves as robots and CI enter our workplaces.

4.2.2.4 Comfort, well-being and acceptance

It is also important to foster **potential end-users' acceptance** of CI human-centered technologies in industrial settings like those under development in AI REGIO. Such an acceptance is critical for their adoption in the workplace in the future and to identify human-related barriers and facilitating factors, ranging from ergonomics, to user experience, comfort, trust, feeling of safety and control over the system and liability allocation, employees' well-being at work On the other hand, the attitude towards CI devices in the workplace is expected to evolve as human experience with machines in general evolves, therefore it is important to be aware of the **generation differences**.

Despite people tend to accept robots quite well and to treat them as social entities, it is still unknown exactly to what level robots and CI solutions will be accepted in the workplace.

Training and education could be useful to overcome the resistance to these new technologies and, therefore, crucial for their acceptance. This is also aligned with the overall need for continuous training inside the factory, which characterizes Industry 5.0 landscape, functional to ensure upskilling and reskilling of the employees in the more and more complex situation in industry,

Moving to another interlinked element, relevant to foster the CI-driven systems' adoption and acceptance in the workplace, is the identification of adequate **well-being metrics** in order to "assess, understand, measure, monitor, safeguard, and improve the well-being impacts of A/IS on humans"[34] [25][35] [26]. Consistent and multidimensional indicators and metrics of success, going beyond traditional ones (which refers, for instance, to profit, productivity, consumption levels, and occupational safety), should be used in this regard. They should be easily understandable and implementable by the developers and designers, as well as easily usable by those affected by the technology: the consideration of these metrics would contribute to avoid unintended consequences and to maximize the benefit from CI innovation, thanks to an adequate evaluation of the impact of CI products, services, or systems on the operators concerned. By prioritizing well-being metrics as an outcome in all CI-empowered system designs, the CI solutions can profoundly increase **human**

[34] IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Wellbeing Committee

[35] IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, "Prioritizing human well-being in the Age of Artificial Intelligence", 2020

flourishing. Among the aspects of well-being which could be relevant for CI system we can mention, for instance: Discrimination & Inclusion, Participation, Engagement, Human Rights, Physical Health, Psychological Health, Satisfaction, Time Balance, Workplace Environment

4.2.2.5 Liability

As previously mentioned, the closer co-working models framed by the introduction of CI-empowered AI solutions into industry challenges existing liability frameworks, making the **attribution of liability blurred**, such as i) in the case of harm arising from the actions of a robot in a **hybrid team of humans and robots working together**, with consequences going far beyond that permitted by the operator/owner of the system, ii) as well as in case of **decisions taken by the machine thanks to its self-learning capabilities** and which deviate from what was initially planned by the producers and expected by the users. The existing schemes of strict product liability for damage arising from defective products seems not suitable for evolving CI systems autonomously learning from their own variable experiences (and of other systems connected to them) and interacting with their context in a unique and unforeseeable manner.

The requirement for human oversight in the context of AI self-learning systems from its design and throughout the entire lifecycle, is not currently foreseen by the existing Union product safety legislation: however, it could be helpful for better safeguarding users.

Further consideration regarding the liability challenges can be retrieved in Sect. 3.1.

4.2.2.6 Respect of human rights and avoidance of algorithmic biases

It must be ensured that the CI-system does not infringe human rights. As regards hardware, it is important that the design process of the device (such as a wereable device like an exoskeleton), can be used by both men and women. As regards software, it is paramount to ensure to avoid algorithmic biases, for instance towards groups or races. This also applies to datasets, in case of data-driven models. For instance they have contain both male and female participants, with different anthropomorphic structures and have to take into account individual differences and providing models respecting such differences, without privileging one group.

4.2.2.7 Accountability, Explainability and Automated decision-making

CI systems must be accountable, including in the real-time decisions about their actions and interactions with humans. Accountable software algorithms must be used, with an excellent level of documentation. The use of black-box components should be minimized and always justified and documented (including answers to questions like: Is the black-box component strictly necessary? What does it do?). The use of learning algorithms should be documented, including their limits.

On the other hand, the personalization of services, especially if relying on the automation of decision-making based on pervasive predictive models, could become ethically problematic.

It can be argued that the underlying algorithms are inherently value-laden and that in case algorithms relieve humans from the responsibility of decision-making, such responsibility should be taken by the designer of the algorithm, including for the ethical consequences of their decisions. On these aspects we have already provided an overview of main positions and challenges in Sect. 3.1 and Sect. 4.2.2.5.

There might be a tension between, on the one hand, the use of CI and its capabilities for achieving flexibility and customization and, on the other hand, the possible disempowerment and inequalities which could occur in case choices are related to profiling inherent in recommendation systems.

4.2.2.8 Safety and Security

It is critical to consider safety risks and to prevent accidents: their possible causes and hazards in human-robot co-existence in manufacturing should be investigated for this purpose. Safeguards and measures capable to avoid their occurrence or, at least, to mitigate their impact should be identified and implemented, so that to significantly reduce accident or hazard probability or accident severity.

Hazards that can cause the robot to collide with, or catch, a human co-worker, eventually injuring co-workers, include both I) impact hazards, caused by events like unexpected movements, failing equipment (e.g. valves, cables, electronics, programs), dangerous work pieces released in an unexpected or erroneous manner; manipulation with hazardous side effects (such as robots bumping into humans), and ii) trapping hazards, where the workers are trapped between a robot and a static object, such as a machine, while the robot is active

Also ensuring security of the devices is a critical aspect and some security issues have direct safety implications, such as an attack that can gain control over the physical actions of a system could cause the device to behave in a physically dangerous fashion on the co-worker. Some others regard the leakage of personal data.

4.3 Human Factors and ergonomics management: an aspect of the Collaborative Intelligence interactions

The goal of human factors is to reduce human error, increase productivity, and enhance safety and comfort with a specific focus on the interaction between the human and the thing of interest. Under these conditions, HF can be seen as an aspect of the Collaborative Intelligence interactions described in section and in particular of the "machines assist humans" channel.

In order to enhance the HF, it is needful to monitor the interactions that occur between machines and workers to identify any problems in their collaborative phases. At the same time, it can be useful manage (i.e., configure, and orchestrate the interactions) that occur between machines and workers, by balancing between machine and human components. Afterward, stakeholders and other users can have the need to simulate these interactions (also when the processes are running) in order to select the most efficient solution that satisfies the case study's requirements. Thus, the idea conceived in AI REGIO is that stakeholders must be supported by a Collaborative Intelligence platform which must enable both contexts where there is a physical interaction between machines and workers, and contexts where there is no physical interaction and instead the interaction between machines and workers is based on voice or visual interaction.

Through this platform, stakeholders can:

- Access to models representing Human and AI processes and in particular HF and their behavious;
- Design the orchestration of human-centred processes workflow in terms of process management and Human-AI interaction;
- Promote harmonization and orchestration between machines and the HF, especially considering the cognitive and physical workload related to manufacturing operations;
- Design the workflow of human-in-the-loop solution through their combination into business processes;
- Use of an orchestration and enactment service to support the design of operational and interaction workflows and the configuration of specific applications and services;
- Check the efficiency of each designed process through off-line and run-time simulations. For this reason, a quantitative modelling of the process is used in order to simulate and to assess

all the possible scenarios. In this regard, some HF can be used as KPI (e.g., Flexibility, Speed, Scale, Decision-Making, Personalization);

- Enable the man's HF in the digital twin loop to simulate and monitor the orchestration of the processes;
- Allow comparison of different solutions of orchestrations in order to select the most efficient solution that satisfies the requirements of a specific scenario in terms of HF.

5 AI REGIO TERESA Framework

This section outlines the **background reference framework for TERESA's development**, in particularly focusing on its concept in AI REGIO and the envisaged respective roles for Didactic Factories and Competent Authorities, besides providing the template to be used as a planning tool by each of the Didactic Factories which will host one TERESA: an overview of the TERESA, outlining the key topics, such as abstract, motivation, Competent Authority to be engaged and CI innovation at stake, will be followed by the elaboration of a detailed Testing Plan, where the main elements and parameters of the TERESA are described, including enabling technologies, use cases to be tested, main objectives, risks and safeguards to be taken.

The design and deployment of the TERESA are based on the **strong interrelation between T7.1 and T7.2**: the Didactic Factories in the project's ecosystem will conceive, design and implement AI REGIO TERESAs, if opportune with the help of supporting technological partners from AI REGIO Consortium, and through the involvement of both the Competent Authority/ies and participants who will voluntarily take part to these testing activities.

5.1 TERESA in AI REGIO: concept and methodology

TERESAs stands for "Technical and Regulatory Sandboxes" for AI.

The Consortium, through its network of Didactic experimental facilities, closely associated to VANGUARD Pilot Plants and Regional / National Industry 4.0 initiatives, in the next stage of the project will develop the TERESAs. Their planning and implementation will rely on a "**hands-on" bottom-up approach** tailored to the specific needs of manufacturing. In this way, AI REGIO TERESA will enable a direct testing environment for innovative CI-empowered products and services aimed at addressing ethical challenges and shortcomings of the regulatory framework concerning such products and services, such those mentioned in Sect. 4.2. A selection of innovative AI applications/tools/services for CI-driven human-machine interaction will be tested/experimented throughout scenarios able to fully point out their complexity in the DFs' facilities, running **experiments on a limited scale and in a secure and controlled way**, according to the "test before invest" paradigm (**technical sandbox**). The experiments will involve volunteers (representing the Civil Society dimension within the sandbox) to test such solutions in real regulatory conditions in a gradual way before going to the Industrial plants, pursuant to a specific testing plan agreed and monitored by the competent authority (**regulatory sandboxes**). This will allow to better understand the relevant regulatory and ethical issues and to better assess the viability of such innovative tools, in particular in terms of their application of and compliance with regulatory, ethical and supervisory requirements. On the other hand, this scheme is also expected to improve the competent authorities' understanding of such innovative AI artefacts, including their opportunities, risks and related regulatory treatment, as well as of their impact on EU policies and values (such as non-discriminatory treatment, human control, autonomy and self-determination).

The **engagement and knowledge-exchange between the Competent Authorities** (regulators, supervisors, policy-makers, innovation agencies, VI representatives, regional or local authorities, etc.) **and DFs** within this controlled environment for safely testing innovation will support AI REGIO innovators in navigating through the regulatory framework and develop a much better **understanding of CA's expectations and legal constraints**. This is expected to allow them to develop human-machine interaction products and services in a regulation-compliant way from the design stage, thus avoiding potential legal risks later, whilst bringing the potential for "reducing the time-to-market cycle" for such products/services.

The TERESA will also deepen the knowledge of how better regulation practices and tools can foster the development and adoption of beneficial CI innovation, contributing to **identifying strengths and weaknesses of existing regulatory policy approaches** to tackle with the regulatory challenges posed by human-machine collaborative environment, in view of highlighting opportunities and possible areas for improvement. In this regard, the adequateness of some provisions of the regulatory reforms under development, in particular the AI Act, could be explored as well.

This mechanism can support to the advancement or further **development of regulations capable of keeping up with the fast pace of innovation**, including some degree of regulatory oversight and support and a facilitated collaboration and communication in a "safe space".

At the same time, TERESAs are expected to concretely contribute to the **operationalization of the ethical principles for trustworthy AI** elicited by the Ethics Guidelines for Trustworthy AI [28] and fine-tuned by the Assessment List of Trustworthy Artificial Intelligence (ALTAI) [29], both elaborated by the High-Level Expert Group on Artificial Intelligence (AI HLEG).

5.2 Didactic Factory Definition & Didactic Factories' Role

This paragraph will illustrate and explain the role of Didactic Factories within the AI REGIO Project and why it's a suited environment for TERESAs. It will introduce the concept of Didactic Factories, how it is defined within the AI REGIO consortium, and what role they will play in service of TERESAs. The following working definition has been defined and used within the AI REGIO Project: ''A Didactic Factory (DF) is an open testing and experimentation facility which extends the services of a Learning Factory towards the materialization of the EDIH "test before invest" pillar. By providing access to technical expertise and experimentation as well as the possibility to "test before invest", A Didactic Factory, like a EDIH, helps companies innovating their business or production.'' The following sub-sections will explain how this definition came to be and what the role of a Didactic Factory is in service of a TERESA.

5.2.1 Finding a Working Definition for Didactic Factories

To set the stage for collaboration within the AI REGIO consortium, it was necessary to find a working definition of a 'Didactic Factory'. During preliminary meetings between partners it became clear that there was a lot of confusion around the topic of Didactic Factories. Furthermore, looking at other European initiatives there are a lot of concepts being used and mixed together, for example: TEFs, DIHs, EDIHs, Learning Factories, Etc. To avoid further confusion on the topic, the following paragraph will illustrate how the project came to the current working definition and what current working definition is being used. This has be done through three methods: 'Research', 'Collaboration & Communication', and 'The Golden Circle' method.

5.2.1.1 Research

In the last decade, the concept of Learning Factories or Didactic Factories started to arise and spread, especially in Europe, both in industry as well as in academia to support education, training and innovation in manufacturing[36] [30].

[36] Abele, E., Metternich, J., Tisch, M., Chryssolouris, G., Sihn, W., ElMaraghy, H., Ranz, F., 2015. Learning factories for research, education, and training. Procedia CIRP, 32, 1-6. doi:10.1016/j.procir.2015.02.187

Some authors[37] [30] gave a clear description of the basic concept of Learning Factories. They state that a Learning Factory addresses both words of the term: It should include elements of learning (and teaching) as well as a production environment. The term learning as opposed to teaching emphasizes the importance of learning by doing which leads to greater retention than following lectures. Learning factories provide a reality-conform production environment for students to learn in. They further made a distinction between Learning Factories in the narrow and Learning Factories in the broader sense. The latter are further away from reality and less hands-on but offer advantages regarding the scalability and location independence. In a Learning Factory in the narrow sense students and factory personnel can learn about the production of a physical product on-site. In a Learning Factory in the broader sense, the product can be a service, or the representation of the product can be virtual.

Didactic Factories take the services of Learning Factories and extend this with services related to drive innovation. As such, the purpose of Didactic Factories is mainly twofold: competency development of students and factory personnel and driving innovation. To allow driving innovation Didactic Factories can also function as a bridge between the Manufacturing SMEs and universities or technology providers.

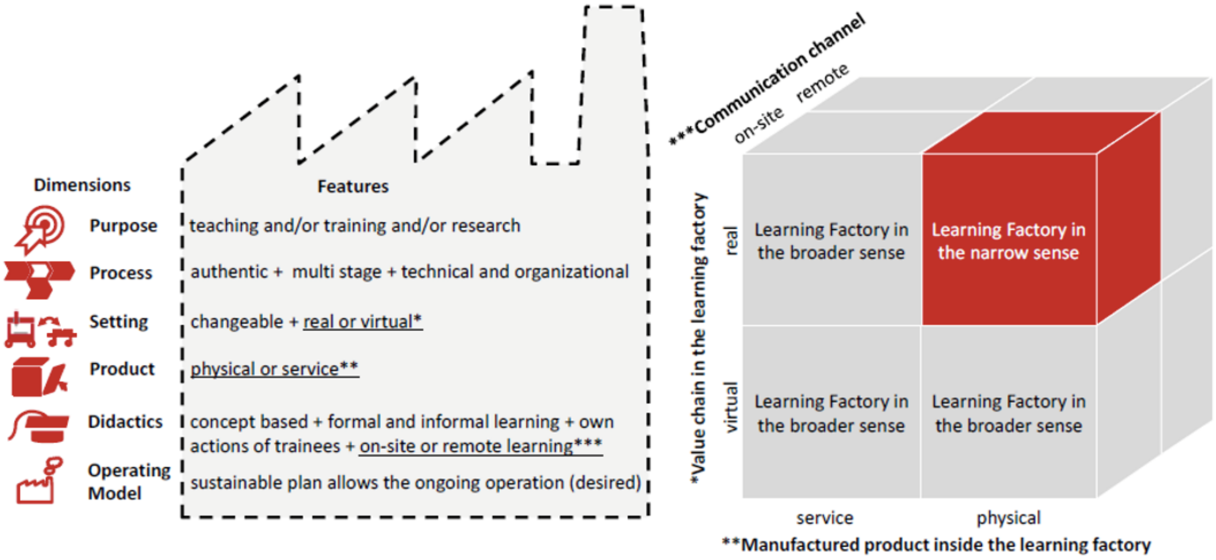
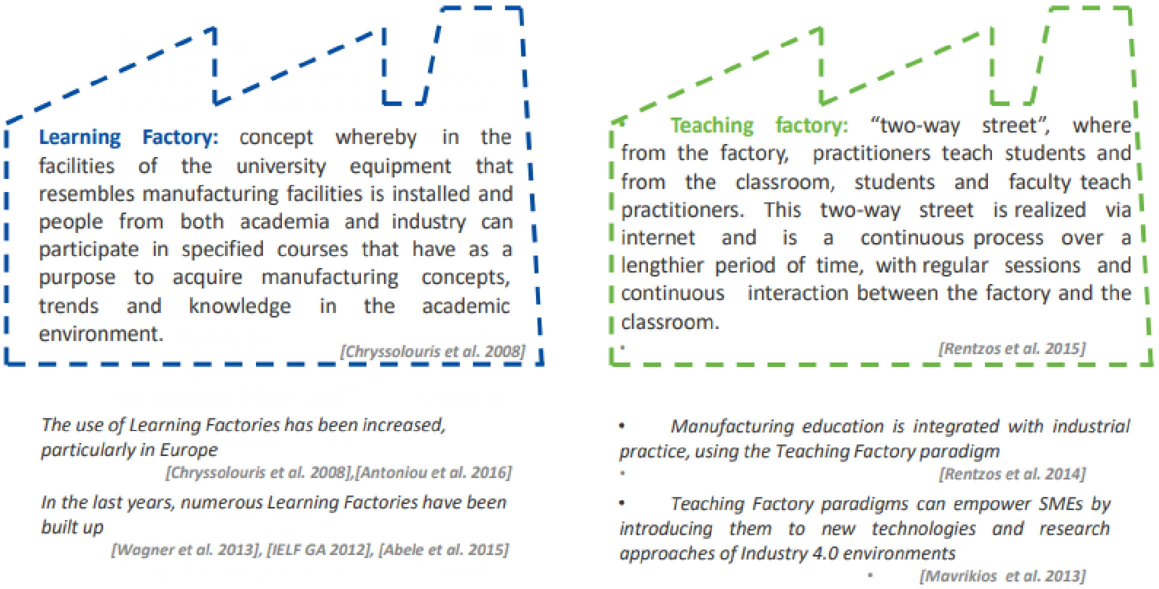

Dimensions
Features
Purpose
teaching and/or training and/or research
Process
authentic + multi stage + technical and organizational
Setting
changeable + real or virtual*
Product
physical or service**
Didactics
concept based + formal and informal learning + own actions of trainees + on-site or remote learning***
Operating Model
sustainable plan allows the ongoing operation (desired)
***Communication channel
on-site
remote
*Value chain in the learning factory
real
virtual
Learning Factory in the broader sense
Learning Factory in the narrow sense
Learning Factory in the broader sense
Learning Factory in the broader sense
service
physical
**Manufactured product inside the learning factory

Figure 2. Key figures of learning factories in the narrow (red cube) and in the broader sense (all grey fields)

[37] Abele, E., Metternich, J., Tisch, M., Chryssolouris, G., Sihn, W., ElMaraghy, H., Ranz, F., 2015. Learning factories for research, education, and training. Procedia CIRP, 32, 1-6. doi:10.1016/j.procir.2015.02.187



Learning Factory: concept whereby in the facilities of the university equipment that resembles manufacturing facilities is installed and people from both academia and industry can participate in specified courses that have as a purpose to acquire manufacturing concepts, trends and knowledge in the academic environment.

[Chryssolouris et al. 2008]

The use of Learning Factories has been increased, particularly in Europe

[Chryssolouris et al. 2008],[Antoniou et al. 2016]

In the last years, numerous Learning Factories have been built up

[Wagner et al. 2013], [IELF GA 2012], [Abele et al. 2015]

Teaching factory: "two-way street", where from the factory, practitioners teach students and from the classroom, students and faculty teach practitioners. This two-way street is realized via internet and is a continuous process over a lengthier period of time, with regular sessions and continuous interaction between the factory and the classroom.

[Rentzos et al. 2015]

- *Manufacturing education is integrated with industrial practice, using the Teaching Factory paradigm*

[Rentzos et al. 2014]

- *Teaching Factory paradigms can empower SMEs by introducing them to new technologies and research approaches of Industry 4.0 environments*

[Mavrikios et al. 2013]

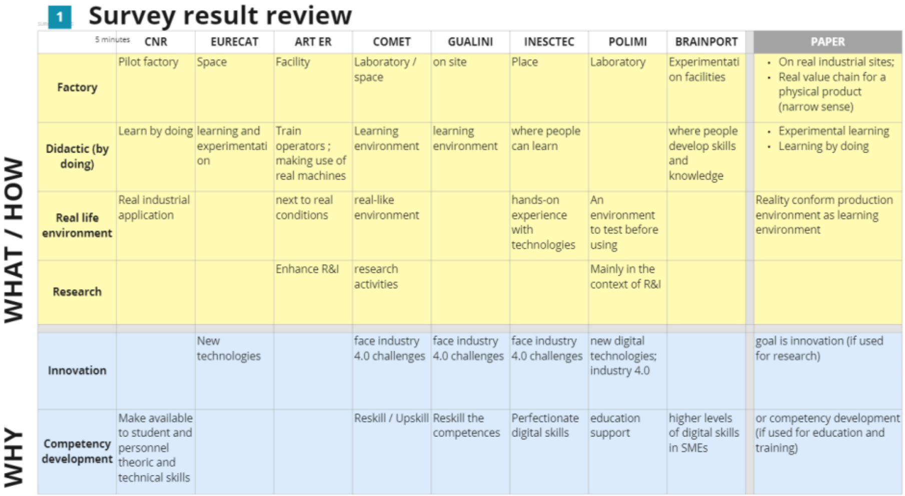
Figure 3. Learning Factory and Teaching Factory

Like Learning Factories, Didactic Factories can be both organized around a physical product or around a service, and the learning can take place on-site or virtually. However, due to the experimentation aspect, Didactic Factories are more likely to be on-site (and hence of the narrow kind). The didactic factory typically offers means and spaces to companies to allow experiments with new products or production methods. Thus, these companies can test before they invest and run lower risks in their innovation activities.

5.2.1.2 Collaboration & Communication

As stated in 5.2.1, there was lack of clarity around the topic of Didactic Factories inside and outside the AI REGIO Consortium. Therefore, the following question had to be asked: *''How do partners and other members perceive a Didactic Factory?''* With these insights it helps understand the role and position of a DF in a business context.

The input for this question has been gathered through: Extensive communication between partners of AI REGIO, Surveys and Workshops.



1 Survey result review

	5 minutes	CNR	EURECAT	ART ER	COMET	GUALINI	INESCTEC	POLIMI	BRAINPORT	PAPER
WHAT / HOW	**Factory**	Pilot factory	Space	Facility	Laboratory / space	on site	Place	Laboratory	Experimentation facilities	• On real industrial sites; • Real value chain for a physical product (narrow sense)
	Didactic (by doing)	Learn by doing	learning and experimentation	Train operators ; making use of real machines	Learning environment	learning environment	where people can learn		where people develop skills and knowledge	• Experimental learning • Learning by doing
	Real life environment	Real industrial application		next to real conditions	real-like environment		hands-on experience with technologies	An environment to test before using		Reality conform production environment as learning environment
	Research			Enhance R&I	research activities			Mainly in the context of R&I		
WHY	**Innovation**		New technologies		face industry 4.0 challenges	face industry 4.0 challenges	face industry 4.0 challenges	new digital technologies; industry 4.0		goal is innovation (if used for research)
	Competency development	Make available to student and personnel theoric and technical skills			Reskill / Upskill	Reskill the competences	Perfectionate digital skills	education support	higher levels of digital skills in SMEs	or competency development (if used for education and training)

Figure 4. Overview of the survey results

We also used other examples of prominent learning examples like the Competence Centre Kaiserslautern in Germany.

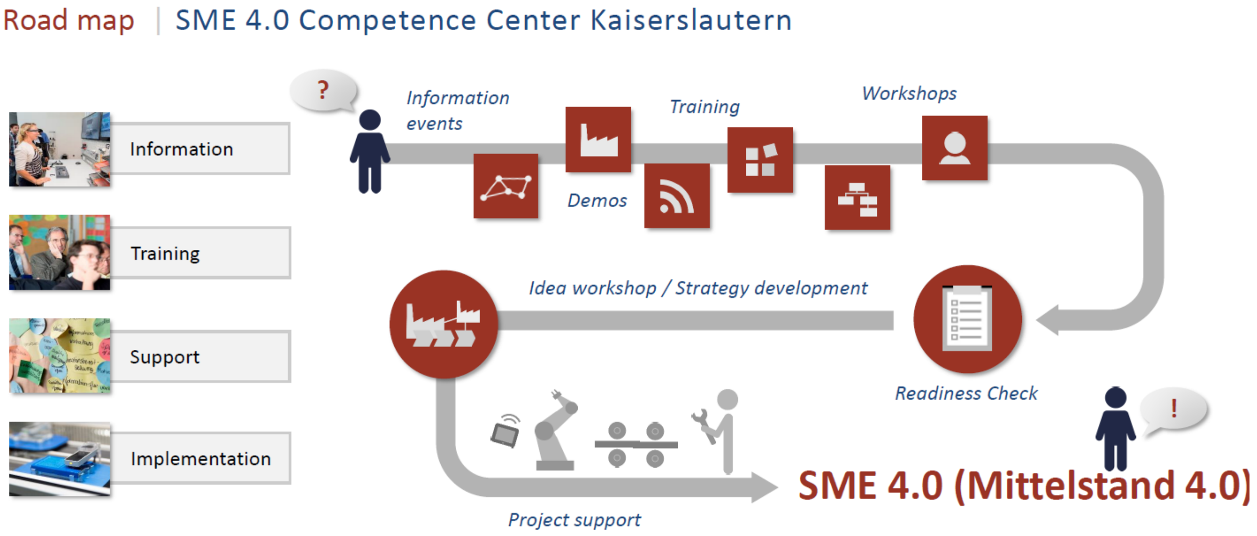
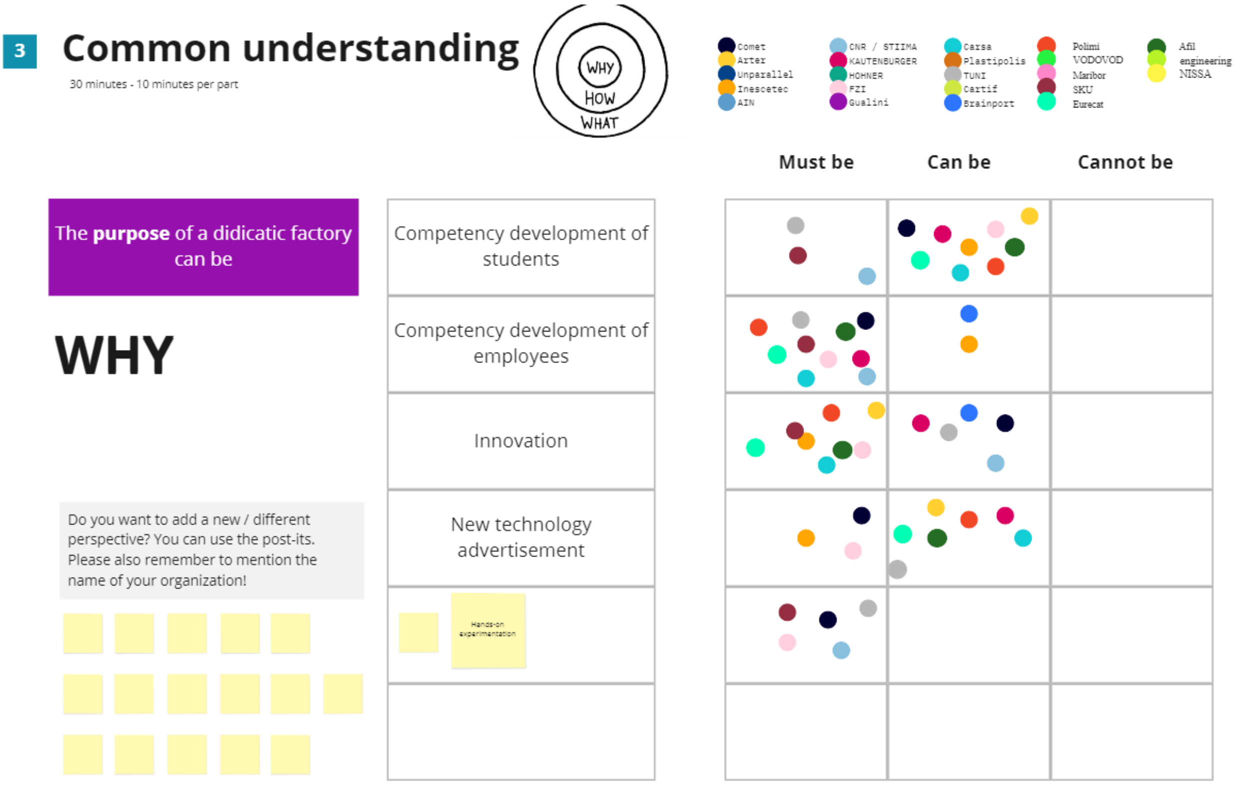

Road map | SME 4.0 Competence Center Kaiserslautern
Information
Training
Support
Implementation
?
Information events
Demos
Training
Workshops
Idea workshop / Strategy development
Readiness Check
!
SME 4.0 (Mittelstand 4.0)
Project support

Figure 5. SME 4.0 Competent Center Kaiseslautern

5.2.1.3 The Golden Circle Method

With the 'Golden Circle' method we have looked at the purpose of a Didactic Factory within the AI REGIO project. Commonly used in marketing practices, this method helps to truly understand the value proposition of a product or service. In this case, a Didactic Factory and its purpose. Thanks to an interactive session with the AI REGIO DIHs and other partners, the following dimensions of a Didactic Factory have been identified:

Why: A DF aims at developing practitioners' competencies and introducing innovation, and also at providing training to students. DFs also function as a bridge between the Manufacturing SMEs and universities or technology providers. The figure below shows the result of the interactive session in detail.



3 **Common understanding**

30 minutes - 10 minutes per part

● Comet	● CNR / STIIMA	● Carsa	● Polimi	● Afil
● Arter	● KAUTENBURGER	● Plastipolis	● VODOVOD	● engineering
● Unparallel	● HOHNER	● TUNI	● Maribor	● NISSA
● Inescetec	● FZI	● Cartif	● SKU	
● AIN	● Gualini	● Brainport	● Eurecat	

The **purpose** of a didicatic factory can be

WHY

	Must be	Can be	Cannot be
Competency development of students	●●●	●●●●●●●●●	
Competency development of employees	●●●●●●●●●●	●●	
Innovation	●●●●●●●●	●●●●●	
New technology advertisement	●●●	●●●●●●●	
Hands-on experimentation	●●●●●		

Do you want to add a new / different perspective? You can use the post-its. Please also remember to mention the name of your organization!

Figure 6. "WHY dimension of a DF"

How: A DF is based on a physical location and might provide virtual tours of its premises and equipment. A DF consists of elements of learning (and teaching) as well as a reality-conform production environment. The figure below shows the result of the interactive session in detail.

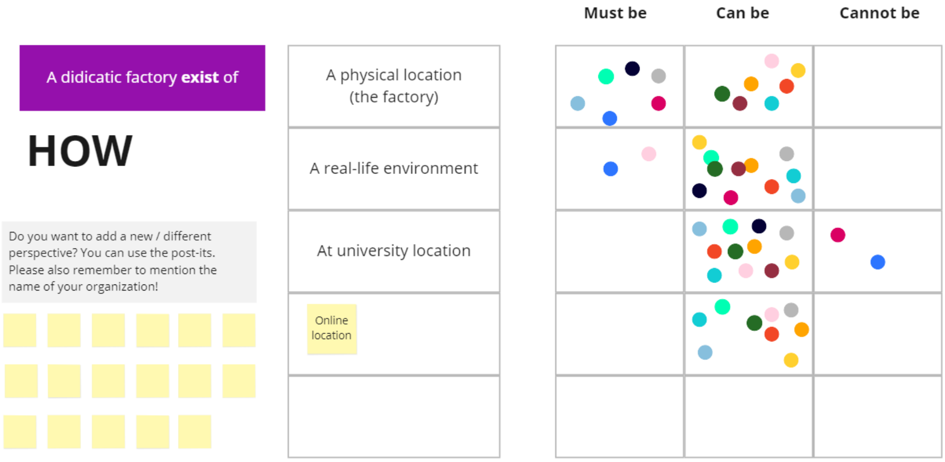

Must be
Can be
Cannot be
A didicatic factory exist of
HOW
Do you want to add a new / different perspective? You can use the post-its. Please also remember to mention the name of your organization!
A physical location (the factory)
A real-life environment
At university location
Online location

Figure 7. HOW dimension of a DF

What: A DF offers training and education by 'learning by doing' and class-room methods, and with its hardware and software assets can perform testing and experimentations. The figure below shows the result of the interactive session in detail. The figure below shows the result of the interactive session in detail.

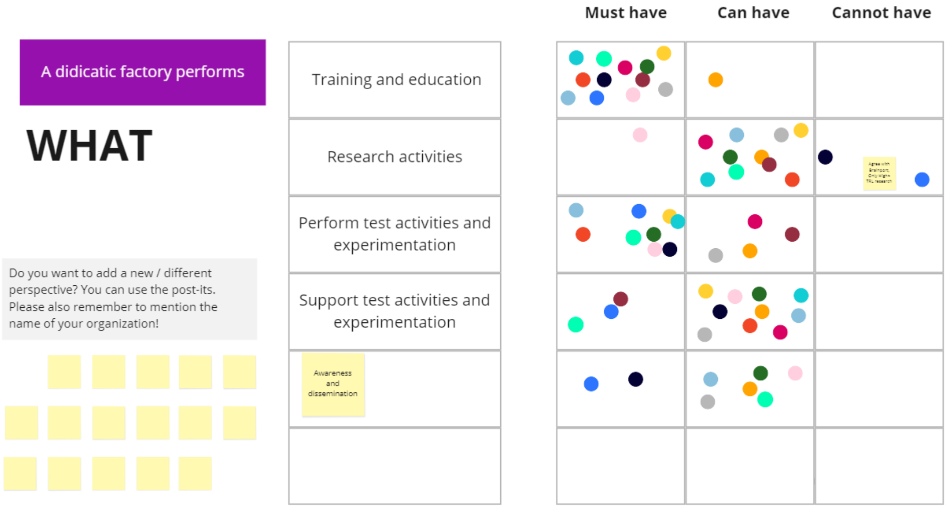
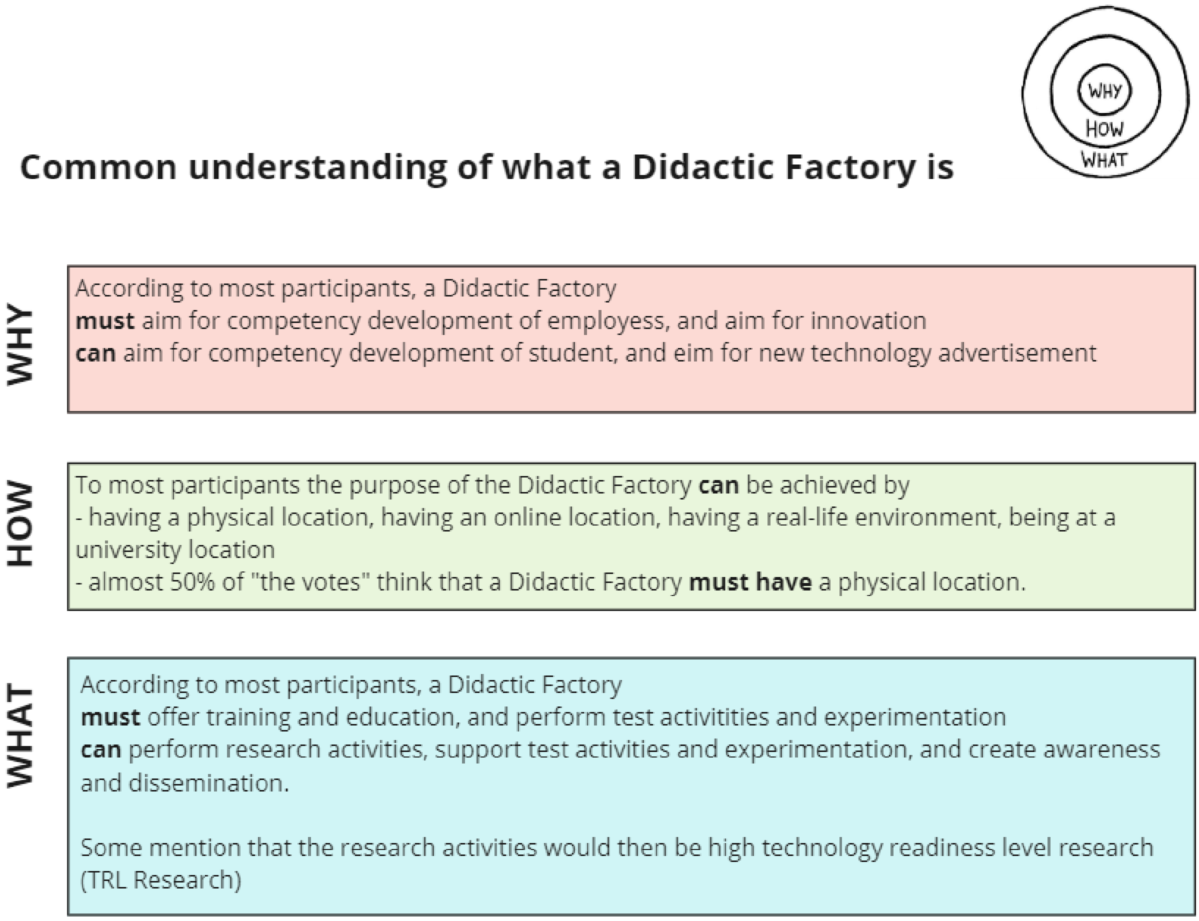

Must have
Can have
Cannot have
A didicatic factory performs
WHAT
Do you want to add a new / different perspective? You can use the post-its. Please also remember to mention the name of your organization!
Training and education
Research activities
Perform test activities and experimentation
Support test activities and experimentation
Awareness and dissemination

Figure 8. "WHAT dimension of a DF"

The final outcomes of the interactive session on what a Didactic Factory is are reported in the figure below.



Common understanding of what a Didactic Factory is

According to most participants, a Didactic Factory
must aim for competency development of employess, and aim for innovation
can aim for competency development of student, and eim for new technology advertisement

To most participants the purpose of the Didactic Factory **can** be achieved by
- having a physical location, having an online location, having a real-life environment, being at a university location
- almost 50% of "the votes" think that a Didactic Factory **must have** a physical location.

According to most participants, a Didactic Factory
must offer training and education, and perform test activitities and experimentation
can perform research activities, support test activities and experimentation, and create awareness and dissemination.

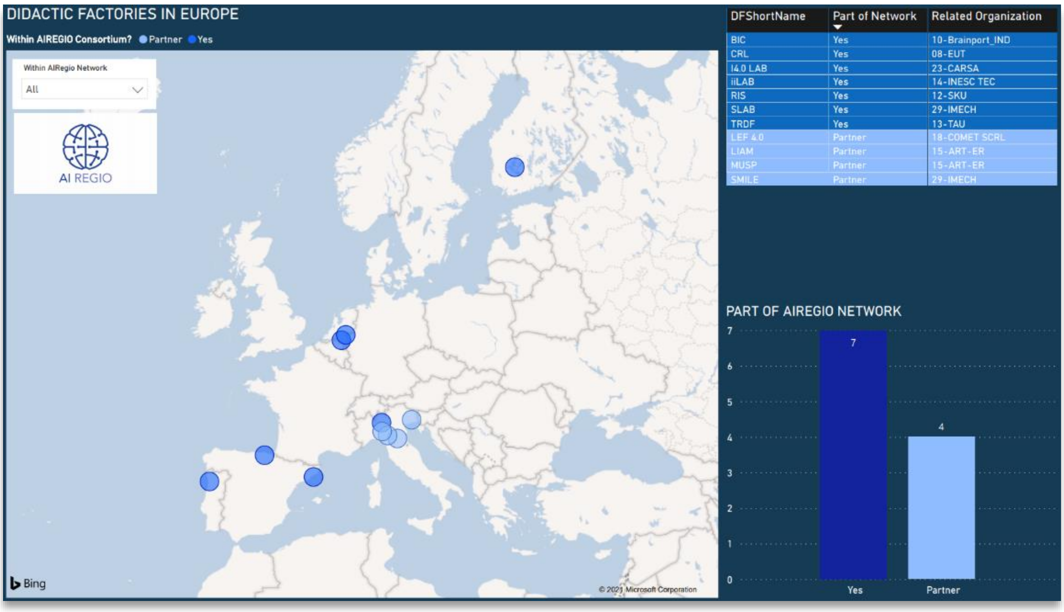
Some mention that the research activities would then be high technology readiness level research (TRL Research)

Figure 9. "Common understanding on the concept of DF"

5.2.2 The Didactic Factories' Role

Within AI REGIO and in the related ecosystem, a number of Didactic Factories (DFs) have been identified. These DFs have two main purposes: support competency development of manufacturing employees and support innovation efforts. To achieve these goals, the Didactic Factories offer training and education as well as facilities to perform tests and experimentation. As it is easy to imagine, within AI REGIO, the focus for didactic factories is on Industry 4.0, digital transformation and AI[38] [30]. Within AI REGIO the aim is to create an network of Didactic Factories to support each other in these previously mentioned ambitions. The following 11 Didactic Factories have been identified as the first wave of ´*AI REGIO Regional Champion Didactic Factories*´:

[38] Mavrikios D, Papakostas N, Mourtzis D, Chryssolouris G (2013) On industrial learning and training for the factories of the future. J Intell Manuf 24(3):473–85



DIDACTIC FACTORIES IN EUROPE

Within AIREGIO Consortium? ● Partner ● Yes

Within AIRegio Network: All

DFShortName	Part of Network	Related Organization
BIC	Yes	10-Brainport_IND
CRL	Yes	08-EUT
I4.0 LAB	Yes	23-CARSA
iiLAB	Yes	14-INESC TEC
RIS	Yes	12-SKU
SLAB	Yes	29-IMECH
TRDF	Yes	13-TAU
LEF 4.0	Partner	18-COMET SCRL
LIAM	Partner	15-ART-ER
MUSP	Partner	15-ART-ER
SMILE	Partner	29-IMECH

Figure 10. Geographical distribution of the first wave of ´AI REGIO Regional Champion Didactic Factories

DF Name	AI REGIO Beneficiary	Location
Brainport Industries Campus	10-Brainport_IND	Eindhoven, The Netherlands
Radboud Industrial Sustainability	12-SKU	Nijmegen, The Netherlands
LIAM Lab	15-ART-ER	Bologna, Italy
MUSP Lab	15-ART-ER	Piacenza, Italy
Lean Experience Factory	18-COMET SCRL	San Vito al Tagliamento, Italy
Cognitive Robotics Laboratory	8-EUT	Barcelona, Spain
Tampere Robotics Didactic Factory	13-TAU	Tampere, Finland
I4.0 LAB	23-CARSA	Bilbao, Spain
SMILE	29-IMECH	Parma, Italy
Industry and Innovation Lab	14-INESC TEC	Porto, Portugal
Smart Lab	29-IMECH	Bergamo, Italy

Figure 11. List of the first wave of ´AI REGIO Regional Champion Didactic Factories

As already highlighted, because of their already functional experimental facilities and their ties to industry problems, SMEs and manufacturing employees, **Didactic Factories represent the ideal place within AI REGIO to implement the Technology-Regulatory sandboxes (TERESA)** where experiments can be run in facilities and situations close to real factories and where constraints can be relaxed which cannot be relaxed in a real plant, while at the same time guaranteeing the safety of the participants involved in the experiment. The increasing digitization and use of AI within factories may pose significant ethical and regulatory issues that are currently not addressed or covered by regulation. The Didactic Factories in AI REGIO, have a crucial role in being able to address these ethical and regulatory issues arising from adoption of AI and human-machine interaction. To support this process, Didactic Factories will:

- identify islands where they can run a **TERESA to experiment Human-AI interaction** in a sandbox. Didactic Factories are free to define experiments that have relevance to them and their partners;
- create and implement a **TERESA experiment;**
- involve and align with the relevant competent authorities, gather the **necessary authorization** that allows exceptions to the legislation, as well as **ethics approvals** or other authorizations if necessary;
- involve, where necessary **technology or industry partners;**

They will start by filling in the TERESA Form described below in Section 5.4: it will be key to plan operations.

5.3 The competent authorities' role

The **TERESAs** are expected to contribute to **enhance the regulatory response to innovation** in the manufacturing domain. In fact, one of the main expected benefits of running TERESAs is the enhancement of the Competent Authority (CA)'s understanding of the new challenges and risks brought by human-machine collaboration and AI in this area, in view of facilitating an adequate policy response. AI REGIO TERESA will enhance the **CA's understanding** about the way in which the regulatory and ethical framework applies to innovative AI products or services characterized by human-machine interaction. In fact, through the dialogue and experience sharing with DFs, the CA is facilitated in its gathering valuable information and understanding of new or altered risks brought by human-machine interaction. The **expected enhanced knowledge gained by CA** can be exploited to support the **timely update of regulatory and supervisory policies**, addressing inadvertent practical barriers to innovation, as well as to reshape CA's monitoring approaches and policies to address any risks to employees, consumers, and market integrity. In other words, the enhanced visibility of broader CI technological developments can also help CA, including supervisors, to **bridge the information gaps** emerging at the edge of the regulatory perimeter, thanks to an improved intelligence and necessary elements to assess whether such regulatory perimeter needs to be adjusted. Furthermore, in this sort of regulatory experimentation, **the CA can closely monitor the actual development and testing of an innovative CI service, product or business model.** If opportune, it can also, on the one hand, **test a certain customized regulatory/ethical approach** to it, in order to avoid regulating it potentially prematurely or inadequately, or, on the other hand, propose **modifications/recommendations** for the advance of the current applicable regime based on the volunteers' feedback.

At the same time, the early-stage liaison by the CA with innovators might encourage them to **engage with CA** in a proactive way.

The TERESA can help DFs and innovators in **reducing ethical and regulatory uncertainties**, lowering the regulatory and ethical barriers to test innovations and bring them into the market. Questions might be asked to better understand CA's expectations and encountered difficulties, to seek clarifications or non-binding guidance about AI-related issues in the context of compliance with the regulatory framework, as well as licencing or registration requirements.

This CA's engagement will provide a **strong signal to innovators about the propensity of the CA to support innovation** and consequently could lead to innovation boost in that specific domain.

The **main roles of the CA** can therefore be summarized as follows:

- Better **understanding of innovation**, by identifying, through an agile learning process, the key barriers that were faced during the implementation and testing period, also in view of defining strategies for risk management for such innovations;
- Help in **reducing regulatory uncertainty** which can be a constrain to the rollout of innovative products and solutions;
- Inform policy making and development through experimentation and learning, identifying areas to **improve the regulatory frameworks**, including soft-law tools such as codes of conduct. This also implies the **contribution to the update regulations** that otherwise might hinder beneficial innovation in relation to AI and human-machine interaction in particular, or at least having the flexibility to amend or revise the process or the general guidelines. In case the involved CA has limited authority to conduct and perform required regulatory reforms, it is advisable that they have a position that could properly **advocate for such reforms**;
- Concrete **signal of commitment** to boost innovation
- **Promotion of engagement** with innovators (but also with volunteers), giving frank and fast **feedback, guidance and indications** on regulatory and ethical requirements and risks and **monitoring** the introduction of new services and products based on participants' experience in order to ensure their **safety and ethically-soundness**. This requires that the CA involves **skilled human resources**, able to evaluate the complex innovations as presented by DFs.
- **Effective consultation and communication with DFs and participants** at the different stages of the TERESA establishment and implementation, thereby helping to **build trust**, facilitate viable partnerships and set appropriate expectations at all levels.

AI REGIO Consortium is also exploring the opportunity to involve with the role of Competent Authority in one or more TERESA a Standardization Body (SB), in particular one or more SBs of the **CEN National Standardization Bodies** (NSBs), which cover 34 countries. The development and valorization of the role of an innovation ecosystem based on the synergy between DFs (in their role of drivers, innovation facilitator), Innovators (AI REGIO Consortium) and Standardization Bodies could contribute to implement the AI Act overcoming some potential barriers and to execute its provisions (for instance in terms of valorizing the role of the DIH and the regulatory sandboxes), besides for moving forward towards the operationalization of the requirements of the ethics guidelines for trustworthy AI (EC's priority) in the manufacturing sector. This **DFs/SB AI REGIO Model for TERESA** could be then introduced, if opportune, in contexts like the CEN-CENELEC Focus Group on AI and replicated in other sectors.

5.4 TERESA Form

The Consortium elaborated the **Form to be filled in by each of the Didactic Factories hosting one TERESA** (one form for each TERESA), which is reported in Annex 1. The form consists of three parts:

- Table for the identification of the **Responsible** for the TERESA concerned;
- **Overview of the TERESA**, where the DF in charge will describe key topics, such as abstract, motivation, Competent Authority to be engaged and CI innovation at stake;

- **Testing Plan**, where the main elements and parameters of the TERESA are described, including enabling technologies, use cases to be tested, main objectives, risks and safeguards to be taken.

6 Plans for the next period

In order to move forward in the creation of TERESAs and in the growth of the Didactic Factory Network of AI REGIO three main steps have been identified.

First, there is a direct need to increase the current participants of the network. For this purpose, the AI REGIO Project will direct efforts towards creating more interest and awareness around its current network of Didactic Factories, including its network's efforts and aims. The AI REGIO Didactic Factories network starts from 11 Champion sites which will be expanded as much as possible inside and outside AI REGIO consortium, especially covering next EDIH and AI TEFs for Manufacturing in any EU Area (e.g. DIH WORLD, DT-ICT-03, AI4EU DIH4AI, DIHNET, EITM).

Secondly, in order to better analyze and categorize the current Didactic Factories and their services AI REGIO aims to analyze every Didactic Factory using the DR BEST Methodology created by POLIMI. With this method, we will assess every Didactic Factory and its services in six dimensions. Unlike the DBEST Methodology, the R dimension for Remotization has been added. During these first months we have identified that most Didactic Factories offer 'Remote Services'.

The third and final step would be to identify and launch TERESA experimentations in the participating Didactic Factories, testing the CI-driven solutions in a real Didactic Factory environment.

Therefore, the following three steps incremental approach is proposed:

1. **Awareness creation**, with several dissemination events (every 3-6 months);
2. **DR BEST collaborative and participative workshops** as well as online self-assessments of AS-IS and TO-BE Service Portfolios (including the R dimension)
3. **TERESA experimentations** identified and launched, testing the CI-empowered human-macgine interaction services, tools and solution in the real DFs' experimentation facilities/environment.

Additionally, we want to check two major assumptions made during these first months:

- **Major assumption I:** can we assume that AI REGIO DFs are a subset of DIHs with some specific **common patterns** in Service Provision?
- **Major assumption II:** can we identify characteristics of DFs as a blueprint for our concept of AI REGIO DFs? (Remote services is the first characteristic we want to put to the test)

7 Conclusions

In this document the key aspects of the **Trustworthy Framework for AI REGIO** project and socio-technical system have been deepened, as a starting point under which to develop project's TERESAs.

The main **legal and ethical challenges** relevant to AI REGIO development and future uptake have been investigated. These include, for instance, the **liability** issues and related uncertainties related to the identification of "who controls whom" and "who controls what" and, thereby, who is accountable and for what. The main complexities derive from the fact that the human behaviour is mediated through an autonomous system, as well as from the existence and interplay of several actors in the "collaborative manufacturing supply chains".

Other challenges, which might affect the well-functioning and competitive positioning of the European Manufacturing sector, arise. They are related to **data ownership and data sovereignty** concerns for the fair and secure sharing, accessing or (re-)using of third party data. Focusing on the legal perspective and the legal basis of the ownership claims, many questions come at stake and the scholars propose diversified solutions, ranging from the need to rely on the protection given by the Confidential information/Trade secrets, to the reference to Copyright in Data, till the need to enact a "sui generis" right for data. Currently, the practitioners are relying on **contract law**: individual contracts cover data ownership, exchange, access to and use of data among the actors along the value chain. Experts commonly agreed that, in the lack of suitable legal instruments, these agreements are capable to manage the data ownership and the control of access and (re-) use of data.

Other legal and ethical dilemmas relevant to AI REGIO, and especially to the human-machine-interaction technologies and human-centric aspects of AI-based manufacturing systems, have been deepened. They include those regarding **Data Protection and Privacy**, the **risk of stigmatization and social sorting**, the concerns for the **safety** of the workers and, in general, the need to prioritize employees' **comfort and well-being** in the CI-driven working environment, both from a physical and from a psychological perspective. From this point of view, the risk of "technostress" and, especially, the risk of development of **emotional attachments to machines** should be prevented.

The document depicts the core elements of the AI REGIO Human-centric CI Approach, and its human-to-machine and the machine-to-human elements, relying on the achievements reported in D5.1 and with an in-depth investigation of the **twofold nature of the HF**, which comprises a technical perspective, as well as a legal, regulatory, ethical and societal viewpoint. Both the "humans in the loop" train-explain-sustain and amplify-interact-embody paradigms, which can be respectively interpreted from the metaphor of a Parent-Child family relationship and of a Caregiver-Elderly relationship, have been addressed. The focus, notably, was on their capability to foster data-and-human-oriented SME digital transformation, in alignment with AI REGIO trajectories towards Industry 5.0 paradigm. By addressing the twofold nature of HF AI REGIO will contribute to conceive and deploy its AI-driven autonomous systems/services, capable of efficiently and effectively interact with Humans according to the value-driven Collaborative Intelligence paradigm. In this perspective, the document outlines how the project is dealing with the human-centric aspects of AI-based manufacturing systems and enabling an immersive AI-based digital workplace, lingering both on technical and on an ELS aspects of these enhanced interactions between information, processes, machines, and people.

On the other hand, besides addressing the topic from a theoretical and methodological perspective, AI REGIO is planning and organizing the work for experimenting these concepts and insights in its TERESA, which stands for **"Technical and Regulatory Sandboxes"** for AI. In fact, AI REGIO, through its network of Didactic experimental facilities, closely associated to VANGUARD Pilot Plants and Regional / National Industry 4.0 initiatives, is setting the ground for the development of these TERESAs relying on a "hands-on" bottom-up approach tailored to the specific needs of manufacturing. They are expected to become a **powerful testing environment for innovative CI-empowered products and services for human-machine interaction**, with the aim of addressing ethical challenges and shortcomings of the regulatory framework concerning such products and services (regulatory sandboxes).

In some of DFs' facilities, identified within the AI REGIO Didactic Factory Network under development, these experiments will be run on a limited scale and in a secure and controlled way, according to the **"test before invest" paradigm** (technical sandbox), pursuant to a specific testing plan. Such a plan will be agreed and monitored by the competent authority, ensuring their improved understanding of the CI tools to be tested and an increased legal certainty of the innovators: they will be able to develop their solutions in a regulation-compliant way from the design stage, thus "reducing the time-to-market cycle" for such products/services. The **DFs/SB AI REGIO Model for**

TERESA, based on the engagement of the Standardization Bodies as possible competent authority within the TERESA, will be explored. This model might contribute to implement the AI Act and to overcome some of its potential challenges and execute its provisions (for instance in terms of valorizing the role of the DIH and the regulatory sandboxes) in the manufacturing sector and beyond.

8 References

[1] D. C. D. Gusmeroli Sergio, «BDVA White Paper - Big Data challenges in Smart Manufacturing Industry,» 2020.

[2] EC, *COM (2020) 65 final - White Paper on Artificial Intelligence – A European Approach to Excellence and Trust,* 2020.

[3] EC, *COM(2020) 64 final - Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics,* 2020.

[4] EP, *Resolution on a civil liability regime for artificial intelligence, (2020/2014 (INL),* 2020.

[5] EP, «Resolution on a framework of ethical aspects of artificial intelligence, robotics and related technologies,» 2020.

[6] J. Villasenor, «"Products Liability law as a way to address AI harms", report from The Brookings Institution's Artificial Intelligence and Emerging Technology (AIET) Initiative, AI Governance series,» 2019.

[7] E. Parliament, *Resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)),* 2020.

[8] E. Commission, *COM/2021/206 final, Proposal for a Regulation laying down harmonized rules on Artificial Intelligence (ARTIFICIAL INTELLIGENCE ACT),* 2021.

[9] E. Commission, *Inception Impact Assessment on the Initiative "Adapting liability rules to the digital age and circular economy",* 2021.

[10] https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai. [Online].

[11] E. Commission, *COM(2019)168) "Building Trust in Human Centric Artificial Intelligence",* 2019.

[12] D. C. H. G. a. o. Martina Barbero, «"Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability. Study prepared for the European Commission by Deloitte,» 2017.

[13] T. Scassa, «"Data Ownership", CIGI Papers No. 187,» 2018.

[14] I. -. I. D. S. Association, «"Data Sovereignty – Critical Success Factor for the Manufacturing Industry",» 2021.

[15] I. D. S. A. -. IDSA, «White Paper "Sharing data while keeping data ownership. The potential of IDS for the data economy",» 2018.

[16] IDSA, *DIN SPEC 27070,* 2021.

[17] E. Commission, *"BACKGROUND DOCUMENT – Workshop on recommendations and impact validation - Technological and economic analysis of industry agreements in current and future digital value chains",* 2021.

[18] E. Commission, *"Introduction to the Final webinar & DRAFT Executive Summary - Technological and economic analysis of industry agreements in current and future digital value chains,* 2021.

[19] G. M. S. Modoni, «AI REGIO D5.1 Collaborative Intelligence and Industry 5.0,» 2021.

[20] H. J. W. a. P. R. Daugherty, «Collaborative Intelligence: Human and AI are Joining Forces,» *Harvard Business Review,* 2018.

[21] M.-P. Pacaux-Lemoine, «"HUMAN-MACHINE COOPERATION: Adaptability of shared functions between Humans and Machines - Design and evaluation aspects",» 2020.

[22] M. S. Gordon Briggs, «"How Robots Can Affect Human Behavior: Investigating the Effects of Robotic Displays of Protest and Distress",» 2019.

[23] D. T. Marie-Pierre Pacaux-Lemoine, «"Ethical risks of human-machine symbiosis in Industry 4.0: insights from the human-machine cooperation approach",» 2019.

[24] A. L. M. A. I. S. Maurice P, «"Ethical and social considerations for the introduction of human-centered technologies at work",» in *2018 IEEE workshop on advanced robotics and its social impacts (ARSO)*, 2018.

[25] M. M. A. d. Graaf, « "An Ethical Evaluation of Human–Robot Relationships",» 2016.

[26] *IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Wellbeing Committee.*

[27] I. G. I. o. E. o. A. a. I. Systems, *"Prioritizing human well-being in the Age of Artificial Intelligence",* 2020.

[28] I. H. L. E. G. s. u. b. t. E. (HLEG), «"Ethics Guidelines for Trustworthy AI",» 2019.

[29] I. H. L. E. G. s. u. b. t. E. (HLEG), «"The assessment list for Trustworthy Artificial Intelligence (ALTAI) for self-assessment",» 2020.

[30] E. M. J. T. M. C. G. S. W. E. H. R. F. Abele, «Learning factories for research, education, and training. Procedia CIRP, 32, 1-6. doi:10.1016/j.procir.2015.02.187,» 2015.

[31] P. N. M. D. C. G. Mavrikios D, «On industrial learning and training for the factories of the future,» *J Intell Manuf,* vol. 24, pp. 473-485, 2013.

[32] «"Collaborative Intelligence: Humans and AI Are Joining Forces"».*Harvard Business Review.*

[33] European Parliament, "Resolution on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012 (INL)", 2020.

[34] Jesper Mathias Nielsen, Jonas Sveistrup Søgaard, Rasmus Winther Mølbjerg and others, "Data ownership, privacy & collaboration when the lines between the physical and digital worlds are blurring" Report prepared by Deloitte in collaboration with Wiliot, 2020

[35] Kadir Alpaslan Demir, Ebru Caymaz, Meral Elci, "Issues in integrating robots into organizations", 2017.

[36] Gordon Briggs, Matthias Scheutz, "How Robots Can Affect Human Behavior: Investigating the Effects of Robotic Displays of Protest and Distress", 2019.

[37] Jenny Waycott, Cosmin Munteanu and others, "Ethical Encounters in Human-Computer Interaction", 2016.

[38] Marie-Pierre Pacaux-Lemoine, Damien Trentesaux, "Ethical risks of human-machine symbiosis in Industy 4.0: insights from the human-machine cooperation approach", IFAC Conference, 2019.

[39] Kadir Alpaslan Demira, Gözde Dövena, Bülent Sezenb, "Industry 5.0 and Human-Robot Co-working", World Conference on Technology, Innovation and Entrepreneurship (WOCTINE), 2019.

[40] S.R Fletcher and P. Webb, Industrial robot ethics: facing the challenges of human-robot collaboration in future manufacturing systems, in: "A World with Robots: International Conference on Robot Ethics: ICRE 2015", 2017

[39] de Graaf, M. M. " An ethical evaluation of human–robot relationships", in International Journal of Social Robotics, 2016

[40] Mario Gleirscher, Nikita Johnson and others, "Challenges in the Safety-Security Co-Assurance of Collaborative Industrial Robots", 2020

[41] Malik, A.A., Brem, A. " Digital twins for collaborative robots: a case study", 2020

[42] Matthew Studley, Alan Winfield, "ELSA in Industrial Robotics", 2020

[43] Maartje M. A. de Graaf, "An Ethical Evaluation of Human–Robot Relationships", 2016

[44] Alexandros Mountrihas, "Human-Computer interaction: Are there any possible risks involved?", 2018

[45] Marina Cugurra, Sergio Gusmeroli and others, AI REGIO D1.11 "Ethics Assessment and Data Management Plan", 2021

[46] W. Patrick Neumann, Sven Winkelhaus, Eric H. Grosse, Christoph H. Glock, "Industry 4.0 and the human factor – A systems framework and analysis methodology for successful development", in International Journal of Production Economics, 2021

[47] Harley Oliff, Ying Liu, Maneesh Kumar, Michael Williams, "A Framework of integrating Knowledge of Human Factors to Facilitate HMI and Collaboration in Intelligent Manufacturing", 51st CIRP Conference on Manufacturing Systems, 2018

[48] Marina Cugurra, Sergio Gusmeroli and others, AI REGIO D2.1 "Legal and Ethical Requirements and Guidelines", 2020

[49] Joan Cahill, "Embedding Ethics in Human Factors Design & Evaluation Methodologies", International Conference on Human-Computer Interaction, 2020

9 Annex 1. TERESA Form

This Annex contains the form to be filled in by each Didactic Factory hosting a TERESA.

9.1.1 Responsible for the TERESA

Name	
Contact details (phone number, email, etc.)	
Role in the DF	
Short profile	

9.1.2 TERESA overview

Name of your TERESA		
Location/facilities	**Didactic Factory name and description**	Insert the name and a brief description of your DF and its strong points
	Description of the experimentation facilities where TERESA will be conducted	Describe your experimentation facilities
	Pictures	Insert some pictures of your DF and/or of the experimentation facilities
	Activities	Briefly describe the most relevant past and current initiatives/testing activities/excellense performed in your DF
Abstract of your TERESA	Provide a snapshot of your planned TERESA as fine-tuned in the next rows, add any details or extra-information that might be relevant.	
Scope of your TERESA	Identify the objectives of your testing activities	
Motivation	Explain the need of "sandboxing" and identify the legal and ethical challenges to be tackled	
Competent Authority/Regulator	**Name**	Complete name
	Function	Describe its functions, focusing on those relevant to the testing
	Power to exception	Clarify if this CA has to power to provide exception to regulatory provisions relevant

		to your TERESA activities and to what extent. In case not, provide information on your plans on this topic (such as applying for Ethics Approvals).
About your AI innovation to be tested	**Description**	Describe the AI tool/service/solution to be tested
	Expected use	Clarify its expected use of the AI solution in the future (domain, type of activities, etc.)
	Expected benefits of the service/tool/solution	Specify the expected advantages and benefit of the AI solution
	Expected risks	Describe potential risks that such AI solution might arise
Training opportunity (if any)	Clarify if you envisage any opportunity for training activities in relation to the realization of this test. In case yes, provide details.	
Legal Framework and regulatory requirements	Give a description of the legal and ethical framework related to the technology or service situation within the test (regulations and legislations, standards and certifications, sector-specific policies). Also non-binding regulatory sources are relevant.	
Legal and ethical issues at stake	Identify the legal and ethical issues relevant in this test	
HF issues at stake	Identify the HF issues relevant in this test	

9.1.3 The Testing Plan

“Name of your TERESA” Testing Plan		
Objectives	Describe the main objectives of the proposed test, what are the main success and failure factors. You are recommended to align these objectives with Competent Authority’s priorities, in order to establish a more effective engagement and collaboration.	
Use case(s)	Clearly describe the use case(s) that you are proposing to test in your test. Please also clarify the scale of testing (that should be limited).	
Enabling technology	Describe the technology infrastructure and other tools, and the location required for testing.	
Timeline and key milestones	Describe the timeline, including the testing duration, and key milestones of your proposed test	
Volunteers to be involved	Identify the type and number of volunteers you are targeting, and how you intend to reach out to them. Clarify inclusion/exclusion criteria and confirm that informed consent procedures will be applied.	
Approvals	Identify the authorisations/approval to be collected, such as Ethics Committee’s approval.	
Risks	**Identified risks**	Identify the key risks associated to the proposed test (safety, privacy, wellbeing, etc.). Risks might refer to the volunteers, but also, for instance, to

		the confidentiality of business information, data sovereignty, etc.
	Mitigating measures	Identify how you will mitigate such risks
Safeguards	Describe the agreed appropriate measures to be adopted, functional to safeguard the volunteers in relation to the testing activity. Safeguard need to be defined on a case-by-case basis, in order to enable flexibility in setting appropriate protection. Aspects that might be considered include, for instance: health and safety, privacy and data protection, property damages, other possible hazards	
Exit Strategy	The test will be terminated in case of infringement of the testing conditions or in case the CA of the DF is not satisfied with the testing outcomes or in case of discovering of critical flaw(s) or unexpected risks during the testing process and these flaws/risks could not be fixed/resolved during the testing period. Provide and briefly explain the exit plan for these and similar cases. Indicate also what factors (if any), in your opinion, would trigger an exit.	
Involvement of the Competent Authority	**Confirmation**	Confirm that you have already shared this Testing Plan with the Regulator/Competent Authority and it is fine for it
	Engagement and supervision	Describe how you will ensure consultation, engagement and results sharing with the CA, as well as its supervision and monitoring as an ongoing process (workshops, reports, site visits, etc.)
	CA’s follow-up actions	Identify the suggested actions/guidelines which could be implemented by the CA upon the completion of the TERESA such as: i) identification of the legal and regulatory requirements required for deploying the innovation/service at a broader scale; ii) possible actions (to be undertaken by the CA or other relevant authorities/Regulator) to amend relevant laws and/or regulations.
KPIs	Identify the Key Performance Indicators (KPIs) that will be used to determine the success of the test	
External engagement	Identify if further actions and engagement are required by external parties and stakeholders	