About this publication
The publication at hand integrates a series of articles and blog posts prepared by the Horizon 2020 STAR project partners. Most of them are available in the STAR website: https://star-ai.eu/blogs-grid. This article collection sheds light on many different aspects of trusted artificial intelligence for industrial use cases featuring also work and achievements of the STAR project towards safe, secure and ethical AI systems.

About STAR
Artificial intelligence (AI) systems in the manufacturing sector are increasingly replacing human tasks improving the automation of production. These systems need to be safe, trusted and secure, even when operating in dynamic, unstructured and unpredictable environments to be able to react to different situations and security threats. Ensuring safety and reliability of these systems is a key prerequisite for deploying them at scale and for fully leveraging the benefits of AI in manufacturing.

STAR, a joint effort of AI and digital manufacturing experts, deploys standard-based secure, safe, reliable and trusted human centric AI systems. STAR researches, develops, validates and makes available to the AI and Industry 4.0 communities novel technologies that enable AI systems to acquire knowledge in order to take timely and safe decisions in dynamic and unpredictable environments, including: Explainable AI, Active Learning and Simulated Reality for fast, safe and efficient online learning and knowledge acquisition, Human Centric Digital Twins, and Security for AI systems. These technologies are validated in challenging scenarios in manufacturing lines, in the areas of quality management, human-robot collaboration and AI-based agile manufacturing.

STAR is funded under the EU Horizon 2020 programme.

To learn more about the project and its outcomes, please visit: star-ai.eu/.

Disclaimer
The content of this publication reflects only the authors’ views and the European Commission is not responsible for any use that may be made of the information it contains.
# Contents

## Part I:

**H2020 STAR Project Overview**

- H2020 STAR: Leading Edge Research for Trusted Artificial Intelligence in Production Lines 6
- A Glimpse on STAR's Reference Model for Trusted AI Systems 10
- Library of Reference AI Scenarios and Use Cases 12

## Part II:

**Human-AI Systems Collaboration and Human-Centred Manufacturing**

- Human-Centered AI: Advancing Interaction 14
- Humans and AI: meeting the challenge of creating effective synergies in manufacturing 17
- Human-Centred AI: Enabling & Facilitating a Climate for Change 19
- Human digital twins to realise production systems where human and machines complement their capacities to achieve better performance 22
- Towards a solution to realise reusable, scalable and extensible digital twins capable of representing humans in a manufacturing context 23
- Improving Visual Quality Inspections with Human-Supervised Machine Learning and Active Learning Techniques: The Philips Pilot of the STAR Project 25
- Humans and AI in Manufacturing: mission impossible? 26
- Human Centred Artificial Intelligence for Agile Manufacturing 28
- Exploring human – machine collaborations for flexible manufacturing 30
- Using machine learning to optimise manual inspection 32
- Leveraging state-of-the-art AI technologies aiming to increase flexibility in automated quality inspection systems 33
- Towards Industry 5.0 and wearables device adoption in industry 34
- Closing the gap between planning theory and shop floor reality 36
- Capturing the different perspective during an early stage of innovation development 37
- Human in the AI loop and how organisations can assess human-centric AI systems in manufacturing? 39
- Sentiment Analysis and Emotion Detection 41
- Employment of NLP within Manufacturing 42
- Occupational Information Databases: data sources for the capacity building 44
- Extending Factory Digital Twins through Human Characterisation in Asset 46

## Part III:

**Industrial Systems Safety**

- SmartFactory KL e.V Plant Scenarios for the STAR project 48
- Safety Zones 50
- Improving the intelligence of the robot with human activity tracking 51
- Synthetic Data for Better and Safer AI Models 52
- Machine Learning for Robots Fleet Optimisation 53
- Testing Mobile Robots in the SmartFactory KL e.V. Plant 54
- Situation awareness for safe robot-human cohabitation in production lines 58
- The Power of Gesture-Based Control in Factory Setup 59

## Part IV:

**Explainable and Trustworthy AI**

- Explaining the decisions of Artificial Intelligence models in Manufacturing 60
- From traditional AI explanations to Human-centered design in XAI for manufacturing applications 62
- Bias management for AI consistent with Human Values 64
- Active Learning: A powerful tool for Accelerated Acquisition of Knowledge 66
- The “little” challenges within the CHALLENGE 68
- The STAR Auditing Framework for Trustworthy AI 69

## Part V:

**AI Cybersecurity**

- Halfway to Security and Data Governance for AI Systems in Manufacturing 72
- STAR’s Blockchain for Data Provenance and Traceability: Tackling the Challenges of Industrial Data Reliability 74
- Securing Digital Manufacturing Platforms with STAR AI Technology 76
- A Perfect Match: How Secure AI Can Shape the Future of Manufacturing 80

## Part VI:

**AI Ethics and Regulations**

- AI that Truly Makes It Work - The All-Functionals Symbiosis as Main Success Factor 84
- AI Norms and Standardisation – the road ahead 86
- AI Maturity Assessment for Manufacturing Enterprise: state of play 91
- How to Make Human-Centred AI Work; Not Just Function? 93
- Manufacturing Data Spaces for Artificial Intelligence: the EC perspective 96
Part I:

H2020 STAR Project Overview
In recent years, we are witnessing the digital transformation of production lines as part of manufacturers’ transition to the fourth industrial revolution (Industry 4.0). Based on Cyber Physical Systems (CPS) and digital technologies like cloud computing, the Industrial Internet of Things (IIoT) and Artificial Intelligence (AI), Industry 4.0 is enabling flexible production lines and supporting innovative functionalities like mass customisation, predictive maintenance, zero defect manufacturing and digital twins. AI is currently the most disruptive digital enabler of the Industry 4.0 era and enables novel use cases like predictive quality management (Quality 4.0), effective human robot collaboration, agile production, and generative software design. State-of-the-art AI systems in industrial plants operate in rather controlled environments. Nevertheless, AI systems in industrial plants must be safe, trusted, and secure, even when operating in dynamic, unstructured and unpredictable environments. Ensuring the safety and reliability of these systems is a key prerequisite for deploying them at scale and for fully leveraging the benefits of AI in manufacturing. This observation is fully in-line with the guidelines of EU’s Expert Group on AI, which mandate that AI systems are robust from a technical perspective and take into account their social environment.

STAR is a joint effort of AI and digital manufacturing experts towards enabling the deployment of standard-based secure, safe reliable and trusted human centric AI systems in real-life manufacturing environments. STAR researches, develops, validates and make available to the AI and Industry4.0 communities novel technologies that enable AI systems to acquire knowledge in order to take timely and safe decisions in dynamic and unpredictable environments.

Moreover, the project researches technologies that will AI systems to confront sophisticated adversaries and to remain robust against security attacks. In this way STAR’s solutions eliminate security and safety barriers against deploying sophisticated AI systems in production lines. The project’s results will be fully integrated into existing EU-wide Industry 4.0 and AI initiatives (notably EFFRA and AMIEU), as a means of enabling researchers and the European industry to deploy and fully leverage advanced AI solutions in manufacturing lines.

STAR acts as a catalyst for ethical AI deployments in production lines, given that the project’s results are fully aligned to the recently published ethical guidelines of EU’s HLEG on AI. Specifically, STAR will produce technical solutions that boost the safety, robustness, and trustworthiness of systems AI in dynamic, real-life settings, while at the same time exploring the legal implications of a safe and secure AI in prominent manufacturing scenarios.

To address the challenges of ethical, trusted, and secure AI systems in manufacturing, including technology and innovation activities in the following areas:

1) Explainable AI: STAR researches and will provide a library of explainable AI (XAI) techniques for manufacturing use cases such as Quality 4.0 and human robot collaboration. The library will contain algorithms that explain the operation of deep learning systems based on their dominant features (e.g., Deep Learning Important Features (DeepLIFT) and Predicion Difference Analysis techniques), other popular XAI techniques (e.g., Lime), as well as algorithms for explainable robotics. STAR will enable recording and replaying the computations that are associated with decisions as a means of understanding how specific inputs lead to given results. The latter process will also facilitate the transparency and predictability of AI systems in the shop-floor, while at the same time boosting security and safety.

2) Active Learning (AL) and Simulated Reality (SR) for Fast, Safe and Efficient On-Line Learning and Knowledge Acquisition: STAR researches AI systems that operate in dynamic manufacturing environments, while acquiring knowledge in a fast and safe manner. Specifically, STAR researches advanced and efficient forms of Reinforcement Learning (RL), including: (i) Active Learning (AL) approaches that enable robots and other AI systems to query human experts about their next course of action. Such AL actions will be employed in cases where robots and other AI systems have low confidence about what to do next, but also as a means of accelerating acquisition of knowledge; and (ii) Simulated Reality (SR) approaches enhance RL systems with on-line simulations as a means of enabling RL agents to simulate the outcomes of their next action before actually taking it.

3) Human Centric Digital Twins for Simulation of the Human-in-the-Loop: STAR researches and will provide advanced Simulation and Digital Twins solutions for AI-based “human in the loop” processes, including human robot collaboration. The project will provide technologies that simulate human behaviour and the interactions between humans and robots, as a means of detecting safety issues in the AI-based manufacturing process. The simulation and digital twins technologies of the project will address several use cases and problems, including: (i) The detection of safety zones for humans; (ii) The deployment of fleets of Automated Mobile Robots (AMR) based on RL techniques; and (iii) The identification of safety issues through monitoring of the worker’s activities and status (e.g., performance, fatigue) in a given task context. To support such simulations, the project will specify and model the digital image of the human worker, while using it to develop Human Centric Digital Twins.

4) Security for AI systems: STAR will research, implement and validate solutions for securing AI systems in manufacturing, including technologies that address attacks by both the training (i.e., poisoning) and the operational (i.e., evasion) phase of Deep Neural Networks (DNNs). The project’s AI security solutions will boost the robustness of DNNs against adversarial inputs and attempts to contaminate the training datasets. Among other techniques they will leverage the project’s XAI library towards identifying and remediating hacked AI systems. Furthermore, the project will provide a decentralised solution for data reliability, which will ensure the availability of high-quality data for training and operating AI algorithms regardless of the distribution and heterogeneity of the data sources (e.g., ERP, smart objects, automation devices, sensors etc.). The solution will leverage a blockchain infrastructure and will exploit background experiences of the project’s partners. It will boost the reliable storage and management of industrial data and of AI algorithms configurations.

STAR’s focus on the above-listed research areas (i.e., explainable AI (XAI), Active Learning (AL) and Simulated Reality (SR) in manufacturing use cases, human centred digital twins for AI in manufacturing, security and trust for AI systems in manufacturing) places the project at the forefront of the global research in AI in general and in digital manufacturing in particular. The project leverages background projects and results of the partners in the above areas, which ensures research excellence and will enable STAR to stand out from similar research initiatives worldwide.
The H2020 STAR project is developing technologies for trusted Artificial Intelligence (AI) systems in production lines. Specifically, the project develops technologies that ensure:

- **Secure and Reliability for Industrial Data:** STAR ensures that AI systems operate over reliable industrial data based on technological solutions (e.g., data provenance, tampered-proof data management) that alleviate the inherent unreliability of industrial data.

- **Secure and Trusted AI algorithms:** STAR develops AI technologies that secure the operation of the AI systems and algorithms that they comprise. In this direction, the project implements cyber-defence strategies that protect and defend AI systems from malicious security attacks. STAR focuses primarily on defences against cyber-security attacks. Physical security attacks are applicable to some STAR systems (e.g., robotics systems used in the project), yet they are not considered in the scope of the project.

- **Trusted Human AI interactions:** STAR focuses on the implementation of trusted interactions between humans and AI systems. On the one hand, the project ensures that AI systems are transparent and explainable to humans towards boosting their acceptance and adoption. On the other, the project focuses also on safe and trusted interactions between humans and AI systems in scenarios like human robot collaboration.

- **Safe AI systems:** STAR includes research towards ensuring the safety of autonomous AI systems such as mobile robots. It focuses for example on the secure placement and movement of Autonomous Mobile Robots (AMRs) in the context of the plant. These systems fall in the broader scope of the safe operation of autonomous systems.

The project has recently produced an initial version of the architecture of its platform, which illustrates the structuring principles and building blocks of a STAR-compliant trusted AI system, which exhibits the above-listed properties. The STAR architecture follows principles of standards-based reference architectures for industrial systems such as the Industrial Internet Reference Architecture (IIRA) of the Industrial Internet Consortium and the Industrial Internet Security Framework (ISIF). In-line with the ISIF, the STAR architecture specifies its security and safety functionalities as a set of cross-cutting functions that are applied to digital manufacturing platforms towards boosting their security and trusted end-to-end i.e., from the devices and the communication end points to the industrial applications. Likewise, in-line with the IIRA, the main functionalities of the STAR platform can be clustered in three main categories or domains according to the IIRA terminology. These three domains are illustrated in the following figure, which provides a high-level reference model for the functionalities of the STAR platform.

Specifically, the three domains are as follows:

- **Cybersecurity Domain:** Comprises functionalities that are destined to ensure the reliability and security of industrial data, as well as of AI algorithms that are trained and operational based on them. The functionalities of these domains support and reinforce the trustworthiness of the project’s functions in the other two domains. This domain includes STAR’s AI cyber-defence strategies, STAR’s blockchain based data provenance techniques, as well as the project’s security risk assessment and security policy management results.

- **(Trusted) Human Robot Collaboration Domain:** Provides functionalities for the trusted collaboration between human and robots. Leverages cybersecurity functionalities, while being used to reinforce functionalities in the safety domain as well. STAR’s simulated reality, active learning and human digital twin systems fall in this domain.

- **Safety Domain:** Ensures the safety of industrial operations, including operations that involve workers and/or automation systems. For instance, functionalities in this domain reinforce worker safety, while catering for the safe operation of AMRs in industrial sites. STAR’s results about the safe placement of mobile robots in industrial plants, the workers’ fatigue monitoring systems, as well as various reinforcement learning techniques fall in this domain.

As illustrated in the figure the functionalities of all domains depend on AI algorithms, including Explainable Artificial Intelligence (XAI) techniques. As such they depend on the STAR AI platform and on the XAI models developed on top of it. XAI plays an instrumental role for the operation of the STAR platform, as it supports defence strategies (in the cybersecurity domain), data generation for simulated reality and active learning functionalities (in the human robot collaboration domain), as well as the development of human digital twins (in the safety domain).

The detailed specification of the STAR reference architecture specifies the building blocks of each one of the above techniques, along with the interactions and interfaces between them. As such it boosts the development of trusted AI systems in general, including STAR compliant systems. For more information about the STAR architecture, you can request our respective deliverable. For more information on the individual technologies and building blocks of the STAR platform, please refer to other blog posts on the STAR web site or to the project’s Open Access Book.
In the context of STAR, a Library of Reference AI Scenarios and Use Cases was created. The AI Scenarios and Use Cases identified are related to manufacturing with an emphasis on scenarios directly related to STAR. To elaborate this library of scenarios the elicitation was performed in two groups. The first one related to the STAR scenarios and use cases and the second, to external and public know scenarios and use cases. The figure below depicts the first steps made for the identification of the scenarios.

The research for external scenarios and use cases, was mainly performed with the help of platforms, as the IoT-Catalogue.com and the EFFRA portal. The IoT-Catalogue.com was utilised to elicit a list of use cases that are known to use AI in manufacturing together with EFFRA which is a portal that promotes the development of new and innovative production technologies provides quite good information. At the same time also the AI4EU project was analysed to elicit the relevant use cases and scenarios targeted by it. Additionally, an analysis was performed on a study conducted by the Centre for Strategy and Evaluation Services (CSES). The name of the study is Opportunities of Artificial Intelligence and provides an assessment of the state of AI adoption in the European industry.

STAR has collected a total of 58 AI Scenarios and Use Cases, distributed in the following way:
- STAR: 10;
- IoT-Catalogue.com: 14;
- EFFRA portal: 18;
- AI4EU: 8.
- Opportunities of Artificial Intelligence Study: 8.

The second figure represents the organogram of how the work progressed after the identification of the scenarios.

The work performed to understand the similarities between the STAR project scenarios and the external scenarios, had the following steps:
- Identify and extract the parameters such as Country, project, and year from the elicited scenarios,
- Analyse and classify the collected information and obtain the domain and category of the AI scenario,
- Relate the STAR project scenarios with the external scenarios.

This enabled the identification of which external scenarios related to the STAR scenarios, and the possible relevance in how the external scenarios can be of use for the implementation and deployment of STAR technologies within the STAR pilots.
Part II:
Human-AI Systems Collaboration and Human-Centred Manufacturing
AI, when aligned with human needs and values, fosters continuous interaction to enhance human capabilities. This shift from technology-centered to human-centered AI signifies a pivotal change in our approach to artificial intelligence. We no longer inhabit a world dominated by technology, but rather a realm where technology harmonises with humanity to enrich our lives. AI is not an omnipotent entity, but a partner, working alongside us to enhance our daily experiences.

Our vision encompasses a future where technology is steered by human values (human-centered), aimed at enhancing our lives, simplifying services, and creating high-quality products. This transformation redraws the relationship between humans and machines, creating a stimulating and mutually beneficial exchange where AI evolves to serve users.

The Human-AI Relationship: a symbiotic and regulated path forward the potential of human-AI collaboration has been realised, particularly with the advent of Large Language Models like GPT and Bard. However, it has also become evident that AI is not yet mature enough to operate without supervision. This underscores the importance of fostering productive human-AI interactions to harness the potential of these new algorithms. The Human-AI Relationship: a symbiotic and regulated path forward the potential of human-AI collaboration has been realised, particularly with the advent of Large Language Models like GPT and Bard. However, it has also become evident that AI is not yet mature enough to operate without supervision. This underscores the importance of fostering productive human-AI interactions to harness the potential of these new algorithms.

The European Union recognises this need and is set to enact the AI Act in 2025, designed to regulate AI and ensure responsible and safe development. The deficiency of AI-enabled systems in "cognitive switching" can be partly mitigated by integrating AI into existing systems. The superiority of human cognitive capabilities over AI in performing across domains is undisputed. Despite massive leaps in the advancement of AI, the supremacy of human capabilities over AI in performing across domains is undisputed. Despite massive leaps in the advancement of AI, the supremacy of human capabilities over AI in performing across domains is undisputed. This goes beyond "transfer learning", aiming to transfer the learned AI capabilities from one original domain to a new one; and it can build on "active learning" approaches, wherein there is an evolving interaction between AI-agents and humans. So why not seek to benefit from integrating rather than eliminating the human from the AI loop? This is not about the human in the AI loop just for the sake of doing so, but for its added value. The concept is illustrated in the figure below.

In recent years, a new era has emerged, where humanity and technology no longer stand in opposition but engage in a synergetic ballet of progress and innovation. The goal is to create solutions where humans and machines engage in continuous dialogue, yielding extraordinary results. This is not merely a human-centered AI approach; it's a journey toward a future where value creation is not only assured but also responsible. This future, AI is not a feared monster but a valuable ally to be comprehended and harnessed.

The debate about the risks associated with deploying AI in industrial workplaces deserves the attention it gets and even more. But it often masks another fundamental one: how human-centric AI can deliver high adding value jobs?

AI-enabled solutions radically reshape the “affordance” of the interaction between human and non-human actors. Affordances are action possibilities provided to an actor by its environment. AI upscales the affordances of both human and non-human (e.g. physical, cyber-physical, or digital entities) actors, operating in production environments. AI-enabled systems can also have vastly superior affordances when effectively integrating the human in the loop. This goes beyond “transfer learning”, aiming to transfer the learned AI capabilities from one original domain to a new one; and it can build on “active learning” approaches, wherein there is an evolving interaction between AI-agents and humans. So why not seek to benefit from integrating rather than eliminating the human from the AI loop? This is not about the human in the AI loop just for the sake of doing so, but for its added value. The concept is illustrated in the figure below.

In recent years, a new era has emerged, where humanity and technology no longer stand in opposition but engage in a synergetic ballet of progress and innovation. The goal is to create solutions where humans and machines engage in continuous dialogue, yielding extraordinary results. This is not merely a human-centered AI approach; it's a journey toward a future where value creation is not only assured but also responsible. This future, AI is not a feared monster but a valuable ally to be comprehended and harnessed.

The Human-Centered AI Approach: a human-centered approach to AI, fostering innovation through close collaboration across various practices, is essential. This approach encompasses refining user experiences, eliminating friction between AI and humans, and addressing technical aspects, including integration with technologies like blockchain. Projects such as STAR, funded by the European Community, exemplify the development of effective AI-human integration solutions. AI’s role becomes pivotal when it is explainable and accessible to everyone. User-explainable Artificial Intelligence empowers non-experts to understand, visualise, suggest, simulate, and interact seamlessly with AI. The Human-Centered AI Approach: a human-centered approach to AI, fostering innovation through close collaboration across various practices, is essential. This approach encompasses refining user experiences, eliminating friction between AI and humans, and addressing technical aspects, including integration with technologies like blockchain. Projects such as STAR, funded by the European Community, exemplify the development of effective AI-human integration solutions. AI’s role becomes pivotal when it is explainable and accessible to everyone. User-explainable Artificial Intelligence empowers non-experts to understand, visualise, suggest, simulate, and interact seamlessly with AI. In recent years, a new era has emerged, where humanity and technology no longer stand in opposition but engage in a synergetic ballet of progress and innovation. The goal is to create solutions where humans and machines engage in continuous dialogue, yielding extraordinary results. This is not merely a human-centered AI approach; it's a journey toward a future where value creation is not only assured but also responsible. In this future, AI is not a feared monster but a valuable ally to be comprehended and harnessed.

The Human-AI Relationship: a symbiotic and regulated path forward the potential of human-AI collaboration has been realised, particularly with the advent of Large Language Models like GPT and Bard. However, it has also become evident that AI is not yet mature enough to operate without supervision. This underscores the importance of fostering productive human-AI interactions to harness the potential of these new algorithms. The Human-AI Relationship: a symbiotic and regulated path forward the potential of human-AI collaboration has been realised, particularly with the advent of Large Language Models like GPT and Bard. However, it has also become evident that AI is not yet mature enough to operate without supervision. This underscores the importance of fostering productive human-AI interactions to harness the potential of these new algorithms. The European Union recognises this need and is set to enact the AI Act in 2025, designed to regulate AI and ensure responsible and safe development. The European Union recognises this need and is set to enact the AI Act in 2025, designed to regulate AI and ensure responsible and safe development. The Human-AI Relationship: a symbiotic and regulated path forward the potential of human-AI collaboration has been realised, particularly with the advent of Large Language Models like GPT and Bard. However, it has also become evident that AI is not yet mature enough to operate without supervision. This underscores the importance of fostering productive human-AI interactions to harness the potential of these new algorithms. The European Union recognises this need and is set to enact the AI Act in 2025, designed to regulate AI and ensure responsible and safe development. The European Union recognises this need and is set to enact the AI Act in 2025, designed to regulate AI and ensure responsible and safe development. The Human-AI Relationship: a symbiotic and regulated path forward the potential of human-AI collaboration has been realised, particularly with the advent of Large Language Models like GPT and Bard. However, it has also become evident that AI is not yet mature enough to operate without supervision. This underscores the importance of fostering productive human-AI interactions to harness the potential of these new algorithms. The European Union recognises this need and is set to enact the AI Act in 2025, designed to regulate AI and ensure responsible and safe development. The Human-AI Relationship: a symbiotic and regulated path forward the potential of human-AI collaboration has been realised, particularly with the advent of Large Language Models like GPT and Bard. However, it has also become evident that AI is not yet mature enough to operate without supervision. This underscores the importance of fostering productive human-AI interactions to harness the potential of these new algorithms. The European Union recognises this need and is set to enact the AI Act in 2025, designed to regulate AI and ensure responsible and safe development. The Human-AI Relationship: a symbiotic and regulated path forward the potential of human-AI collaboration has been realised, particularly with the advent of Large Language Models like GPT and Bard. However, it has also become evident that AI is not yet mature enough to operate without supervision. This underscores the importance of fostering productive human-AI interactions to harness the potential of these new algorithms. The European Union recognises this need and is set to enact the AI Act in 2025, designed to regulate AI and ensure responsible and safe development. The Human-AI Relationship: a symbiotic and regulated path forward the potential of human-AI collaboration has been realised, particularly with the advent of Large Language Models like GPT and Bard. However, it has also become evident that AI is not yet mature enough to operate without supervision. This underscores the importance of fostering productive human-AI interactions to harness the potential of these new algorithms. The European Union recognises this need and is set to enact the AI Act in 2025, designed to regulate AI and ensure responsible and safe development. The Human-AI Relationship: a symbiotic and regulated path forward the potential of human-AI collaboration has been realised, particularly with the advent of Large Language Models like GPT and Bard. However, it has also become evident that AI is not yet mature enough to operate without supervision. This underscores the importance of fostering productive human-AI interactions to harness the potential of these new algorithms. The European Union recognises this need and is set to enact the AI Act in 2025, designed to regulate AI and ensure responsible and safe development. The Human-AI Relationship: a symbiotic and regulated path forward the potential of human-AI collaboration has been realised, particularly with the advent of Large Language Models like GPT and Bard. However, it has also become evident that AI is not yet mature enough to operate without supervision. This underscores the importance of fostering productive human-AI interactions to harness the potential of these new algorithms. The European Union recognises this need and is set to enact the AI Act in 2025, designed to regulate AI and ensure responsible and safe development. The Human-AI Relationship: a symbiotic and regulated path forward the potential of human-AI collaboration has been realised, particularly with the advent of Large Language Models like GPT and Bard. However, it has also become evident that AI is not yet mature enough to operate without supervision. This underscores the importance of fostering productive human-AI interactions to harness the potential of these new algorithms. The European Union recognises this need and is set to enact the AI Act in 2025, designed to regulate AI and ensure responsible and safe development.
Humans and AI: meeting the challenge of creating effective synergies in manufacturing
By: Christos Emmanouilidis and Sabine Waschull / University of Groningen

But how to design-in the incorporation of the human in the AI-loop? An AI-enabled production environment brings outcomes which are not the result of the actions of individual actors, but of their collective activity. Therefore, it is the collective involvement of humans and AI in production activities that needs to be designed-in. In the STAR project we have taken key steps to do just that. By involving a representative range of an AI-enabled ecosystem stakeholders, including technology and AI providers, manufacturers, legal and ethics experts, system integrators, and diverse range of user perspectives, we organised co-creation workshops for each one of the pilot cases. We were not deterred by the pandemic restrictions. We successfully held the workshops using virtual collaboration tools. In each workshop, we had an accelerated collaboration process through virtual collaboration boards. An example is seen here:

1. humans can help/augment the AI,
2. where AI technology can help/augment humans,
3. the joint human-AI activities,
4. expected outcomes and their effects on human workers and work design, as well as operational effects, alongside with success criteria.

Such design co-creation workshops were organised for:
- the Philips case on quality inspection,
- the DFKI case of safe human – mobile robot collaboration in Industry 4.0 enabled environments, and
- the IBER-OLEFF case on agile production planning,

These workshops help the STAR project advance in the next steps towards designing the human-centric AI solutions for the manufacturing lines of the future.

The fear of AI taking over human jobs is one that has to be addressed by creating value – adding human jobs in future manufacturing. We will be more successful in that pursuit if human – AI synergies are designed-in by the involved stakeholders: this is our aim in our STAR agile development process with more co-creation workshops in the next development and testing phases of the project – stay tuned!

The participants first worked to develop user stories; stories were connected with functionalities and components needed for implementing them; there were all linked to outcomes of the envisaged human-centric AI solution, addressing how:

Great Capabilities to Improve
Integrated ecosystems sustain life and provide us with an amazing habitat. People and the ecosystems we live in, in this Digital Age, have great capabilities to improve and sustain the quality of life for all.

As we face and urgently need to deal with many societal challenges, we need a climate for change. Various of such societal challenges can be identified in the domains of manufacturing, supply chains, logistics, maintenance and related industry domains.

As these domains will remain essential parts of our society and economy, a climate for change in these essential parts of our ecosystems is needed as well. Safe, trusted and trustworthy Artificial Intelligence (AI) and other or related knowledge, processes, technologies, human intelligence and experience may be an excellent enabler and facilitator to help cater for and sustain such future-proof ecosystems.

The whole supply ecosystem, including sourcing, engineering, manufacturing, assembling, logistics and the like, as well as the related organisations, professionals, partners and customer involved, and the respective societies, ecology and economy can benefit from access to, use and exchange of data, information, knowledge and experience. Digital platforms, AI, intelligent systems, cognitive (edge and IoT) computing, robotic process automation (RPA), cobots, distributed intelligence and autonomous systems are further expediting this process by connecting, inter-connecting respectively hyper-connecting organisations, individuals, communities, societies and data with tens of billions of objects and entities.

Human-Centred AI: Enabling & Facilitating a Climate for Change
By: Arthur van der Wees / Arthur’s Legal, Strategies & Systems
Where To Start?

What can an entrepreneur, company, sector, community or other groups in manufacturing, industry and related sectors and domains do to create overall positive impact while also having a viable and economically sustainable value model, with related business models and financial and other feasibility models to get things both started, going, trusted, growing, scaling, resilient and future-proof? Having a big vision and focusing on the horizon is important, but having a clear starting point is one of the main prerequisite success factors.

With that in mind, it is recommended to start with identifying and establishing the particular challenge(s) one would like to focus on, for instance by using the 12 Societal Challenges for Future of Living, as visualised below. These are in line with both the vision of the European Commission as well as the United Nations’ Sustainable Development Goals (SDGs). These Societal Challenges are obviously intertwined and interconnected.

Let’s have a closer look to Societal Challenges: Demography respectively Skills & Jobs. Where and why may AI in Industry 5.0 context be valuable, appreciated and even necessary? First some backgrounds:

A. Societal Challenge Nr. 4: Demography

Within the European Union, there is a decline in working-age population. It’s expected to reduce by 13.5 million (or 4%) by 2030 compared to 2018. This, as the EU population size will shrink by 5% between 2019 and 2070, to 424 million inhabitants, by using the 12 Societal Challenges for Future of Living, as visualised below. These are in line with both the vision of the European Commission as well as the United Nations’ Sustainable Development Goals (SDGs). These Societal Challenges are obviously intertwined and interconnected.

B. Societal Challenge Nr. 11: Skills & Jobs

According to the OECD, 65% of the kids in schools today will have jobs that haven’t been invented yet. This indicates that we apparently are not yet sure what the future will look like, but that we do for sure acknowledge society will look very differently in a decade. The World Economic Forum points out that among the top 10 most essential skills of the near future are: analytical thinking, empathy, creativity, reasoning, complex problem-solving, self-management, and technology development and use.

Clearly, this list resembles a more intertwined combination of both the right part of the brain with the left part, than currently commonly seems the case. These two Societal Challenges and backgrounds already demonstrate that AI in Industry 5.0 context may be valuable, appreciated and even necessary to address these societal challenges in industry and related society and economy:

A. When focusing on the Societal Challenge of Demography, combining and deploying innovative processes, data and technologies to augment the capabilities of people, industry, supply side and demand side can be a helpful mechanism to compensate this expected decrease in productivity and levels of welfare and quality of life.

B. When focusing on the Societal Challenge of Skills & Jobs, three questions that come to mind are (i) how will the future of work change the industrial sector, and the looks of our urban and rural societies, (ii) how to keep the veins of trade and human values running through our communities, and (iii) whether technology will displace more jobs in 10 years than it creates, or vice versa. With all these questions raised, what role will and can AI play in combination with human interaction?

With this, the European stakeholders, society and economy can build, deploy, use, enjoy and even export the most trustworthy human-centric AI for Industry 5.0 and related digital (eco)systems and services all over the world. As Commissioner Breton formulates: ‘Europe has everything it takes to lead the technology race’. In our own words: Europe has great capabilities.
Human digital twins to realise production systems where human and machines complement their capacities to achieve better performance

By: Paolo Pedrazzoli, Andrea Bettoni, Elias Montini / SUPSI, SPS Lab

Despite the increasing level of automation, workers’ characteristics, skills, behaviours and psychophysical conditions have a relevant impact on the performance and operations of manufacturing production systems. However, as of today, all these elements are almost neglected in the digital representation of the factory, not allowing to get the most from human and machines interaction.

Humans and automation systems must complement their capacities in order to achieve improved manufacturing performances, thus their historical data, status and evolution must be available for analysis and optimisation. To achieve such a goal, it becomes imperative to create digital representations not only of production systems, but also of workers, considering context data (e.g. assigned job, current workplace, current shift, training program), quasi-static data (e.g. specific worker needs, skills, age) and real-time sensor data (e.g. accelerations, Heart Rate, Galvanic Skin Response, Temperature). All these data can be used to feed:

1) monitoring models and algorithms focusing on specific features and attributes (e.g. detecting workers’ fatigue, estimating mental stress),

2) behavioural modules elaborating the workers’ current status to make predictions and simulate its evolution over time and

3) decision modules identify decisions in order to intervene in the digital and/or in the physical world.

Today, such human characterisation and ontology are only partial and not combined into the factory digital twin. The characterisation of the workers in terms of knowledge, skills, personal needs, intellectual and sensorial capacities and interactions with the factory entities is a cornerstone to actually develop an efficient human-factory relation. The true challenges lie in the fact that 1) humans need far advanced models to support their behavioural paths toward factory enhanced productivity and safety and 2) that these models need a complex understanding of the workers in terms of their features and behaviours.

In STAR, a Human Digital Twin is proposed to address these challenges. The STAR’s Human Digital Twin can be considered as a single source of truth thanks to AI-based intelligent modules, to make them available for machines or human decision-makers and monitoring systems.

Towards a solution to realise reusable, scalable and extensible digital twins capable of representing humans in a manufacturing context

By: Niko Bonomi, Andrea Bettoni, Vincenzo Cutrona, Giuseppe Landolfi, Elias Montini and Paolo Pedrazzoli / SUPSI, SPS lab

Thanks to Industry 4.0, solutions that support monitoring, simulation, optimisation and decision making in manufacturing systems are growing exponentially. Most of these solutions rely on the concept of Digital Twins (DTs). NASA introduced the concept of Digital Twin in 2012 as “an integrated multi-physics, multi-scale, probabilistic simulation of a flying vehicle or system”. Since then, this concept has evolved and been adopted in various fields. Thanks to the technologies introduced by the Industry 4.0 paradigm, DTs have gained tremendous importance in the manufacturing industry. DTs have been successfully used for mirroring and simulation of industrial environments, predictive maintenance, virtual commissioning, anomaly detection and product life cycle optimisation.

More recently, with the advent of the Industry 5.0 paradigm and the reaffirmation of the crucial role of the worker in production systems, DTs must also enable the representation of humans. The Human Digital Twin (HDT) will be a foundational technology to facilitate the integration of human workers in an Industry 4.0 environment, enabling communication, data aggregation, simulation and planning. In the last 5 years, applications for HDTs have emerged in the manufacturing industry, mainly involving worker monitoring, production planning and control, human-robot collaboration, and adaptive automation. However, the implemented models are mostly application-oriented and not reusable.

Currently, there is no solution to support the creation of HDTs, forcing industrial solution architects to resort to ad-hoc implementations and models. On one hand, existing commercial platforms such as Predix or Watson IoT can collect data from a large number of devices and machines but are expensive and require the right support to be integrated in existent production systems. On the other hand, open-source solutions such as Ditto or RAMI AAS, although very interesting for both industry and research, are machine-centric and their information models do not allow to properly model humans.

In STAR, thanks to the close collaboration with the partners of the consortium, the Sustainable Production System Lab of SUPSI is realising a platform to support the development of DTs. The platform enables the representation of humans and contextual entities in the factory, and their interactions. The resulting platform has been obtained thanks to:

---


Towards a solution to realise reusable, scalable and extensible digital twins capable of representing humans in a manufacturing context
By: Niko Bonomi, Andrea Bettoni, Vincenzo Cutrona, Giuseppe Landolfi, Elias Montini and Paolo Pedrazzoli / SUPSI, SPS lab

(a) a specific information model derived by the metamodel proposed in1;
(b) a modular infrastructure to ease the instantiation of HDTs and their components2.

The STAR HDT can be considered as a single source of truth and a central access point to factory entities’ data. The HDT embeds the digital representation of workers, which are seamlessly integrated with the DT of the production system and can be used by AI-based modules for a further operation (e.g., prediction tasks). The first prototype of the platform has been released and will be used to support different applications within the project, including workers’ fatigue estimation, AGV path planning, worker training paths definition. The prototype exploits the MQTT protocol to manage data flows from machines and devices to the HDT, and it provides an orchestration module to help the users in defining and managing their own HDT. Users can manage entities (e.g., workers, machines, sensors) and attach AI-based modules to the HDT by means of this component. The platform provides also a dedicated component to manage data flows history, serving collected data as time series for further analytical and monitoring tasks.

In our previous article, Leveraging State-Of-The-Art AI Technologies Aiming To Increase Flexibility In Automated Quality Inspection Systems, we discussed a specific use-case within the Philips pilot of the STAR project. The use-case is suitable for utilising human-supervised machine learning (AI) for visual quality inspections in the production process. These inspections are critical to ensure the delivery of high-quality products within the Philips factory in Drachten. However, due to the complexity of the inspection process, which includes small anomalies, short cycle times, complex part handling, and broad range of products to be inspected as well as the costs associated with automating these inspections, many of them are now still done manually.

To overcome these challenges, the STAR project envisioned that AI models could decrease the need for manual inspection and improve production speed. At the same time, as AI models become more prevalent in industrial settings, it is imperative for their acceptance and adoption that they work well and can be trusted by its users. This is where we use the term explainable artificial intelligence (XAI). This is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. XAI is very important for humans to be able to trust the outcome of these very complex models. Very often, the data scientists not even fully comprehend why the model makes a certain decision.

To make sure the decisions made by the model are accurate and in line with what a human quality inspector would have decided, the STAR consortium envisioned an AI model based on active learning technology. In active learning, the deployed visual inspection model is improved iteratively through the use of manually revised samples of the streaming data where the model’s confidence is the lowest. In this revision process, the operator can then be provided with defect hinting by using heatmaps to label the selected samples more efficiently and accurately. Various defect hinting techniques are available for this purpose (such as GradCAM and similarity heatmaps) and each have their own effect on the manual revision process.

Apart from its usefulness during data sampling, utilising such explainable AI techniques can make AI systems more interpretable to humans by providing insights into the models’ rationale behind a prediction. While still in the early stages of development, we already have demonstrated promising results and are confident that with utilising such techniques it is possible to automate certain aspects of the inspection process while maintaining high levels of quality control.

---

Humans and AI in Manufacturing: mission impossible?

By: Andrea Ferretto Parodi / GFT Italy

The World Manufacturing Forum 2020 Report says: “Artificial Intelligence is not novel in manufacturing. In the last decade however, thanks in part to advancements in AI algorithms, computational power, connectivity, and data science, it has gained more importance as companies increasingly see it as a driver for competitive advantage. However, the lack of experienced talent to work with AI, lack of know-how, and the need for accurate data remain important challenges for organisations in adopting AI.”

Unlike other more recent and greenfield application domains mostly related to citizens and B2C, the business benefits for AI adoption in Manufacturing needs to be contextualised in the 6Ps dimensions of an Enterprise. How AI-driven Digital Transformation affects company’s Products (e.g. connected cars), Processes (e.g. maintenance), Platforms (e.g. MES ERP PLM systems), People (e.g. professions and skills), Partnerships (e.g. Digital Innovation Hubs and SMEs) and Performance (e.g. twin transition digital-green business indicators).

More specifically, speaking of the People dimension, WMF2020 report says “Currently and even more so in the future, people are consistently interacting with AI. As a result, this changes the nature of work and technological interaction. Particularly, the modern workplace will be affected by the implementation of more AI into processes.” All research and innovation initiatives aiming at introducing AI in manufacturing need to address the challenge of human adoption of increasingly AI-based autonomous systems. New roles and professions, new skills and competencies need to be developed and adopted in the Factory, in the Product Design and Engineering Departments, in the Value Chains.

But how does the Human-Machine interaction happen? How to model it, how to simulate it and how to improve and enhance it in the workplaces? In 2018, Harvard Business Review published an article entitled “Collaborative Intelligence: Humans and AI are joining forces”, envisaging a Human-to-Machine train-explain-sustain and a Machine-to-Human amplify-interact-embody interactive processes. As a matter of fact H-to-M interaction is like a Parent-Child interaction where more experienced and knowledgeable beings are able to train about processes and procedures, to explain why certain decisions have to be taken and to sustain autonomous systems in the accomplishment of ethical, legal, behavioural and governance issues. The M-to-H interaction is similar to the relationship Caregiver-Elderly where more capable beings need to amplify physical or cognitive capabilities of less able people, need to enhance their interaction with other beings or data, need to embody their knowledge and capabilities like in a robotic arm or in an exoskeleton.

According to the EU Commission, Industry 5.0 “attempts to capture the value of new technologies, providing prosperity beyond jobs and growth, while respecting planetary boundaries, and placing the wellbeing of the industry worker at the centre of the production process.” In the Industry 5.0 perspective, next generation AI systems need to consider human factors and human acceptance as a key competitive advantage for their adoption in Manufacturing Industry. Collaborative Intelligence workplace models, simulation and operational systems is one significant way to achieve Industry 5.0 objectives and a more human-centric and sustainable factory of the future.

The STAR ICT-38-2020 project is working in this direction, focusing on leading edge AI technologies with wide applicability in manufacturing environments, including Explainable AI, Active Learning, Simulated Reality, Human-Centric Digital Twins for AI, security and trust for AI systems in manufacturing. The STAR marketplace will be soon available for manufacturing SMEs to grasp and adopt the most recent “AI for Manufacturing” solutions also in the field of Human-AI Collaborative Intelligence. STAY TUNED!!!
Human Centred Artificial Intelligence for Agile Manufacturing

By: Mihail Fontul / IBER-Oleff

The STAR project operates in the IBER (Iber-Oleff SA) value chain with the objective of boosting scientific and technological capacities and competences, and carrying out research and development activities whose results will allow IBER to expand and strengthen its value chain in the fields of products, processes and services, namely:

- Streamlining the use of agile and adaptive manufacturing systems by incorporating advanced online process sensing & monitoring systems and different parameters and variables into the manufacturing process and products, including non-destructive testing techniques, to measure dimensions and defects, with contactless and agile technologies;
- Promoting greater flexibility of the production process through a new and innovative integrated management concept based on the flexibility, versatility and synchronisation of production cells, using dedicated information systems.

A key aspect of the IBER pilot project is manufacturing of a customised product in an agile production system that allows for greater efficiency, speed, operation and maintenance monitoring. The concept of this pilot project is human-centred and will be supported at the production management decision-making level by a future artificial intelligence platform, connected to the existing digital factory. It is expected that the future artificial intelligence platform will allow to accelerate the decisions at the production management level, through simulations of the production processes in real time, reaching the desired flexibility of them. An agile production system will allow rapid reconfiguration of production processes to accommodate sudden customer orders, produce a highly configurable product with quality and in the shortest possible time frame.

Advanced online process sensing & monitoring systems and different parameters and variables into the manufacturing process and products, including non-destructive testing techniques, to measure dimensions and defects, with contactless and agile technologies should be incorporated on these adaptive manufacturing systems. The maximisation of productivity due to the linearisation will have important consequences in the reduction of the work peaks and consequently in the reduction of the number of extra work hours, as well as the reduction of the production down times.

Throughout this process the human factor plays a predominant role, given that our production processes are not fully automated (nor are they intended to be). Nowadays the training of human resources in the various production processes is planned and carried out in advance, in accordance with appropriate operational methods. The flexibility of human resources training, to accommodate the agile production mentioned above, will be achieved with the help of the future artificial intelligence platform, which will contribute, on the one hand, to the improvement of operative methods and consequently the reduction of human errors associated with learning and assimilation, and, on the other hand, due to continuous monitoring of production processes, by reducing human errors associated with overspecialisation, which usually occurs when an operator is performing the same functions in the same workplaces over a given period of time. Practically the phenomenon can be described as the loss of operator concentration due to overconfidence, with consequences on productivity and quality of work performed.

The implementation of the future artificial intelligence platform, fully integrated with other existing digital platforms, will succeed if in terms of confidentiality, integrity, evaluability, non-repudiation and authenticity of transmitted, existing or processed elements in the databases, it is achieved. The artificial intelligence system will have to be developed to identify inappropriate data, eliminate it from processing, so as to prevent the contamination of the learning process itself, leading to wrong analyses and alerts. Inadequate data will be those that have been read by improperly functioning sensors, those that can be injected into the system by humans and are poorly conditioned, or those that are brute force altered by elements outside the company.
Exploring human – machine collaborations for flexible manufacturing

By: Jelle Keizer / Philips

The Philips factory in Drachten, the Netherlands, is an advanced factory for the mass manufacturing of consumer goods (e.g., shavers, OneBlade, baby bottles, soothers) where over 2100 enthusiastic colleagues from all over the world work together to develop and produce products that improve people’s lives. By putting focus on sustainability through innovation, we are always striving to work on the edge of current technical development. However, pushing for innovation is a process that is hard to do alone. Therefore, we are privileged to team up with, among others, RTO’s, technology providers, and other manufacturers in the STAR project, developing knowledge on key enabling technologies and solutions for manufacturing challenges of today and the future.

One of these challenges revolves around the introduction of flexible manufacturing. Within the production of shavers in Drachten there is a strong emphasis on standardisation, automation, and short cycle-times in its production lines. These production lines are often specifically tailored for the mass production of one product series in the most efficient way. However, due to a shift in customer demand, production lines need to be reconfigured more often to be able to produce different products. In the current production setup, these reconfigurations are expensive and time-consuming. During the STAR project, AI-based solutions will be developed to facilitate more flexible production approaches.

As a global leader in the manufacturing of mass-produced consumer products, Philips Drachten supplies a use-case revolving around ‘HUMAN – COBOT Collaboration improving robust Quality Inspections by Vision’ as a pilot demonstrator in a relevant production environment. This pilot offers the opportunity to demonstrate results developed during the STAR project and gives insight into how these results can be leveraged to improve flexibility of today’s as well as future production systems.

To keep a clear vision of the goals, the pilot line within the factory in Drachten explores three different topics regarding the creation of such a flexible production system.

1) Easy reconfiguration for automated part handling,
2) Human supervised learning for visual quality inspections,
3) Safe collaboration between human and cobot.

During the first topic, flexibility from a part handling perspective is explored. This means that research is done into the design of a system that is useable in many different scenarios where an incoming part can be detected, identified, and handled based on the identification of the part.

Next to flexible part handling, also flexible visual quality inspection is critical to ensure the delivery of correct products. Normally, visual quality inspection systems are trained based on extensive datasets and can be easily optimised due to the mass-production of products. However, in low-volume production, these extensive datasets are often not available. Therefore, flexible visual quality inspection systems that can be trained based on small, incomplete datasets, and human input are developed and tested to be able to guarantee the required quality without the need of an extensive dataset.

Finally, the pilot will demonstrate optimal collaboration between humans and machines to achieve flexibility and to make that succeed, safety is critical. Therefore, the third focus point revolves around the creation of a safe production system where human and machine can work in a shared space were unsafe situations are preemptively identified and dealt with. Next to the physical safety of such a collaborative system, also the digital safety and mental well-being of the human in the loop is explored.
Using Machine Learning To Optimize Manual Inspection
By: QLECTOR and JSI

Artificial Intelligence is permeating an increasing number of aspects of our life. It is also increasingly permeating all pores of the manufacturing industry, which enables to achieve more efficient production, and also a friendlier one towards customers and the manufacturing workers.

Quality control is considered one of the most critical manufacturing activities since it ensures the products conform to a set of requirements and specifications, building trust with the customer, boosting customer loyalty, and reinforcing the brand’s reputation. One of the means to realise such quality control is through the visual inspection of manufactured products. Nevertheless, while performed manually, such inspection poses several challenges. First, operators’ decision whether a product is defective or not is subjective, and certain variance exists based on the operators’ training and experience and factors such as their tiredness. Furthermore, such an inspection approach has limited scalability (e.g., it requires training and scaling the number of inspectors proportional to the production scale). On the other side, automated visual inspection guarantees the same criteria for all inspected products while doing so efficiently and with minimal downtimes. Suppose such systems use machine learning to identify potential defects. In that case, a decision must be made up to what confidence level the models’ outcomes are trusted to meet quality policies, deferring the rest of the products to manual inspection. Therefore, a hybrid approach can alleviate much of the human effort required for visual inspection while still relying on humans for more complex cases.

Manual visual inspection can be enhanced with artificial intelligence too. Generative Adversarial Networks can be used to generate synthetic images, to provide a stream of images with an equal part of good and defective products. This can force the operator to pay greater attention than when inspecting an imbalanced stream of images, where most images correspond to non-defective manufactured products. Furthermore, defect hints can be provided to the operators to help them find and label the defects more quickly, increasing the velocity and quality of the throughput.

For further insights on this topic, we invite you to read our paper “Towards a Comprehensive Visual Quality Inspection for Industry 4.0,” which we presented at IFAC MIM 2022 conference. We will be glad to hear your thoughts!

Leveraging state-of-the-art AI technologies aiming to increase flexibility in automated quality inspection systems
By: Jelle Keizer / Philips

In the previous article related to the Philips pilot in the STAR project; Capturing the different perspectives during an early stage of innovation development, we talked about the first stages of developing use-cases where we focused on capturing the different perspective for systems and components to be developed during the project. In this article, we are diving into one of the specific use-cases worked on in the Philips pilot during the STAR project. Specifically, the second use-case, human supervised learning for quality inspections.

Visual quality inspections are critical in the production process to ensure the delivery of correct products. A large portion of (potential) non-conformities within the Philips factory in Drachten are related to the visual appearance of parts and products. Due to the complexity (e.g. short cycle times, complex part handling, high gloss multi curved products) and related costs for (partial) automation, as well as the reconfiguration of related quality inspection systems, a lot of visual quality inspections at the Philips factory in Drachten are still a manual task. The relatively high labor costs for these manual inspections, and the results of rather subjective quality inspections performed by the different quality inspection professionals are not ideal inputs for autonomous process control of our manufacturing processes.

Therefore, solutions are being explored within the STAR project in collaboration with the technical partners aiming to implement a system that will make setting up automated quality inspections easier & faster by applying techniques like active learning. By doing this, the goal is to develop a system that can be implemented within the factory, and that can be used to easily set-up an automated quality control for a new product. The system can learn from the quality inspection professionals on the job, and after receiving enough information about the quality of the products, the system will be able to take over the quality inspection. This way, manual inspections are only performed during the learning phase of the system.

By doing this, collaboration between human and machine is explored aiming to leverage the best of both machine and human where the human provides the machine with the required information after which the machine can take over and the human can focus its efforts on other topics.
Towards Industry 5.0 and wearables device adoption in industry

By: Vincenzo Cutrona, Elias Montini / SUPSI, SPS lab

The digital representation of production systems is getting very relevant in the last decade. Nowadays there are countless examples of machinery and process monitoring solutions to monitor machines, processes and factories, including Digital Twins (DTs). However, the continuously evolving needs of manufacturing end-users require also to represent humans in the digital world, including their intents, behaviours and conditions, realising the so-called Human Digital Twins (HDTs)

Wearable devices (also referred to as wearables) play a fundamental role when HDTs are applied to monitoring applications. Indeed, the raising of advanced, precise, and low-cost sensors on-boarded on wearables enable the collection and processing of human physiological data to support analytics in a variety of applications. Wearables are a category of electronic devices that can be worn as accessories, embedded in clothing, or even implanted in humans’ bodies. Wearables are crucial to enable the Industry 5.0 paradigm, which revolves around the concepts of HDTs and Human-Cyber-Physical systems to facilitate human-machine cooperation.

Different types of wearables are available in the market, including smartwatches, smart clothing, and smart glasses, on-boarding different sensors to fulfill different needs, e.g., fitness assistants, health monitoring, GPS/sport tracking. Bringing wearables to manufacturing production systems opens to several opportunities:

(i) tracking and monitoring operators’ performance, behaviours and conditions;

(ii) supporting operator activities through innovative interfaces and support systems. For example, physiological data from wearables has been used for the estimation of workers’ exertion and mental stress, enabling the adaptation of collaborative robots’ behaviour.

In STAR, SUPSI is developing an AI-based solution to estimate the physical fatigue exertion of workers based on physiological data and operators’ characteristics. Dealing with this goal, one of the challenges faced by the research team was:

"Which is the best wearable to be applied in an industrial context for human fatigue detection?"

The large wearables market makes it difficult to immediately identify the devices best suited for a specific industrial application. This challenge is recurrent in almost all the projects where researchers, industrial experts, and companies embrace I5.0, where the demand for technology selection methods and approaches is increasing. However, common practices and guidelines are still lacking.

As a result of the STAR project, SUPSI’s research team proposed a methodology for selecting wearables suitable for I5.0 applications, which is very easy to adopt thanks to its practicality and business-oriented approach.

The methodology has been presented by Elias Montini at ETFA 2022 in Stuttgart during the special session “Industry 5.0 – Augmenting the Human Worker in Balanced Automation Systems” organised by Tamás Ruppert (University of Pannonia), and David Romero (Tecnológico de Monterrey).
Closings the gap between planning theory and shop floor reality

By: QLECTOR and JSI

Industries nowadays are faced with a number of challenges such as time loss by manually creating and updating production plans and schedules, time loss by fine-tuning schedules via phone or email, inaccessible knowledge locked within planner and team coordinator, number and duration of downtime due to slower response to unplanned events, difficult management of a diverse product portfolio, long and unpredicted lead times, high intermediate inventory, etc.

Advanced in ubiquitous AI can learn to predict lead times more accurately due to external circumstances, with such an accuracy operational downtimes can be prevented or reduced significantly, allocation of human resources can become more optimised, planners and team coordinators can take advantage of automatic planning options and can dedicate more of their time tasks with higher added value, inventory can be lowered and the solution can help approaching towards the goal of lean manufacturing.

Innovative STAR solutions help improve the already cutting-edge Qlector LEAP product (https://qlector.com/solutions.html), which can save 1 day/week for planners and team coordinators, simulate the production process for several day in advance, enable just-in-time delivery, replace repetitive task with added value tasks, reduce lead times, enable higher turnover ratio, enable on-time preparation of tool changes and finally, increase OEE.

Although competitive circumstances in many industries, such as automotive or chemical, already push many production processes to its limits, there are still plenty of possibilities for improvements on an organisational level, which help tackling the challenges mentioned above. Just like Google Maps navigations enables drivers to travel between point A and point B by the most suitable route according to the current situation and predicted road congestions, a production steering solution could guide the production process by its optimal way towards the goal.

Capturing the different perspective during an early stage of innovation development

By: Jelle Keizer / Philips

In the previous article related to the Philips pilot in the STAR project, Exploring human-machine collaborations for flexible manufacturing, we talked about the importance of exploring flexibility in manufacturing, the different use-cases we are developing within the STAR project, and how the expected results might be leveraged in order to improve flexibility and further develop our current production systems to work towards the factory of the future. In this article, we are diving into one of the first stages of the development of the use-cases where we decided to focus on capturing the different perspective for systems and components to be developed during the project.

To clearly define the different use-cases and to get an insight in the different perspectives of relevant stakeholders, we decided to run a workshop in collaboration with our partners in the STAR project. During this workshop we wanted to involve people that could help us define the use-cases and come up with requirements from different perspectives. Therefore, we chose to involve stakeholders from all different stages of the project’s life cycle. We invited research partners, technology developers, domain experts and end-users to all contribute to defining the different use-cases as formulated in the previous blog.

During this workshop a small introduction into the use-cases were given to all participants by introducing the vision and goals behind the use-case and providing a preliminary visualisation by presenting the current processes (as-is scenario) versus the expected future processes (to-be scenario). Subsequently, by using a collaboration tool all participants were instructed to define user-stories for the different use-cases one at the time.

These user stories are a popular method for representing requirements by using a template such as “As a [role], I want [requirements], so that [benefit of requirement]” and by gathering the user stories created by the invited participants we aimed to gather as much input as possible from the different points-of-view. The participants could create user-stories for every user they could imagine, but in the end, we could categorise the different user as the 5 users defined below:

- Organisational user, which focused on the company goals and requirements,
- Technical user, like technical support staff & mechanics,
- Operational user, like production managers, team leads, and operators,
- Technology provider, which focused on requirements during development,
- Researcher, which was focused on the requirements from a continuous improvement point-of-view.

Finally, the different user stories created were categorized and documented in such a way that based on user stories we could define different components functionalities that we need to develop along the course of the project to provide as much value to the end user as possible. In conclusion, we can say that involvement of different types of users in the development process for our use-cases and exploring their point-of-view proved to be very successful and is deemed as a positive contribution to the development of the use-cases and therefore to the project.

In the STAR project we are concerned with key questions about how human-centric AI solutions in manufacturing should be designed, developed, deployed, and importantly, critically evaluated.

We take a co-creative approach to the design, as reported before. We base the development on an appropriate reference architecture. STAR deployment includes specific technology choices reported in our respective deliverable (which can be requested) and our Open Access Book.

We also have concrete examples of how users interact with AI-driven solutions. Here is an example of AI-driven visual quality inspection work implemented by the University of Groningen. An operator in the quality control team (from our IBER OLEFF partner) has the opportunity to interact with an AI-driven system, thanks to a synergy with our partners from the Jozef Stefan Institute (JSI) and Qlector. The user is not only offered post-processing hints or explanations but is additionally provided with prototype images of similar cases, making the process more intuitive. In this way, a job profile in a quality control team of the future might involve more such activities, including data labelling and interaction with AI, and less manual or physical work. This is an example of how job profiles evolve with AI.

But how the industry can evaluate the success of AI-driven industrial systems which are meant to be trusted and human-centric?

A common pitfall is to take a solely technology-driven perspective. This would not suffice for the same reasons that a technology push is never enough for successful innovation. For example, it will not be enough to evaluate success solely based on the accuracy of machine learning models (whichever appropriate nonetheless accuracy definition may be employed).

Another viewpoint is to solely focus on operational performance. There is no denying of the importance.
of such a perspective. But it is unlikely to be enough if it ignores a third viewpoint - that of human factors, which includes also ethics and work design aspects.

The STAR partnership addresses ethics, work design, and societal factors too, for example, https://star.ai.eu/bias-management-ai-consistent-human-values, https://star.ai.eu/how-make-human-centred-ai-work-not-just-function. Yet, the safest and more secure systems might be those that allow nothing to be done. The key message is that it would be wrong to prioritise any of the above category of factors over the others, when designing, testing and evaluating human-centric AI-driven systems.

In the STAR project we put forward a systematic approach to human-centric AI systems evaluation, taking all such factors into consideration. The approach is tested in practice on industrial use cases, and wider communities are engaged to share experience, scrutinise the approach, and learn best practices from each other.

More information on this topic was presented during the STAR Interactive AI Co-Creation Workshop on How to Enable Safe, Secure, and Ethical AI in Manufacturing held at the Innovation Cluster Drachten (Philips site) with the additional support of the AI Hub of North Netherlands.

Understanding how customers like products as well as how technologies are perceived by operators is key for the success of manufacturing products and processes. Sentiment Analysis and Emotion Detection are one of the main exciting features developed within Artificial Intelligent (AI) laboratories that can be exploited for equipping computers and robots with human-like sentiment understanding. These technologies applied to the manufacturing domain can help to quickly detect the satisfaction level of operators and users and, therefore, to promptly make an action either to solve an issue or to support activities.

Modern implementations behind Sentiment Analysis and Emotion Detection are built on top of Transformers, deep learning models that have millions of features and that describe how words are used in speech or text. They can deeply understand how a sentence should be interpreted according to human-like understanding and suggest a score which represents the positiveness or negativeness of human expressions. Also, they can detect these scores for several characteristics of a product that users are talking about. For example, in a review about a product like "My car has a very good engine and brakes, however, the windows are quite small and limit the visibility" transformers might be used for doing the so-called Aspect Sentiment Analysis where first aspects are detected (for example "engine", "brakes", "windows"), and for each of them a sentiment score is associated. For example, "engine" and "brakes" can have a score of 0.8 associated with a transformer indicating that the customer is satisfied, while a score of 0.2 can be assigned to the "windows" to indicate that the customer is not satisfied with them. In doing so, a company can identify which actions have to be taken to improve the underlying product. At scale, companies can analyse, investigate, and explore large amounts of opinions expressed by customers and, hence, be able to evaluate the level of satisfaction of product utilisers. In the long term, detecting and solving issues in the designing and manufacturing of products as well as the identification of pitfalls in manufacturing processes will increase the customers’ satisfaction and will support the building of the customers’ loyalty.

As a consequence, Sentiment Analysis and Emotion Detection are two key technologies to be leveraged within several domains including manufacturing to enhance the companies-customers relationships, thus making a step ahead toward the user satisfaction and win-win situations. Many of the AI systems and components used in STAR require interaction with humans and machines, these interactions can benefit from Sentiment Analysis and Emotion Detection, to improve the communication with the end-users, but also in some internal use cases in which operators and machine collaborate, allowing to detect issues and recommend solutions.
Employment of NLP within Manufacturing

By: Diego Reforgiato Recupero, Nino Cauli and Rubén Alonso / R2M Solution and University of Cagliari

Natural Language Processing (NLP) is a subset of Artificial Intelligence that helps identifying key elements from human instructions, extract relevant information and process them in a manner that machines can understand.

Integrating NLP technologies into the system helps machines understand human language and mimic human behaviour. For example, Amazon’s Echo, Microsoft’s Cortana and Apple’s Siri make extensive use of NLP technologies to interact with the users.

NLP technologies can help in the interaction with the machines and speed up the operation of different types of manufacturing systems, cutting down the response time. Imagine a scenario where a manufacturing company hires a data scientist to collect shopfloor worker information and analyse all the machine readings, reporting any sort of problems. One disadvantage to this scheme is that by the time the management reads the report one problem might have happened causing damage to the entire process. If a computer or robot with sensors and NLP technologies embedded is employed, this might analyse information coming from the machines, reports from customers and information from workers, in order to obtain relevant information about the process. This computer or robot might even communicate with users and accept input in natural language.

Within the manufacturing industry, the NLP might be adopted for example for the following tasks:

- **Process Automation:** The use of NLP technologies in the manufacturing process allows the automatic processing of information in natural language and the execution of repetitive tasks like paperwork and report analysis.

- **Inventory Management:** Analysing data about the stock, sales and user reports of certain products is essential to assess the correct decisions for a company to optimise and maximise profits. By leveraging NLP technologies, the resulting benefits are: 1) the entire process becomes more comprehensive; 2) it is more difficult to incur errors related to the analysis of sales; 3) it is easier to analyse the manufactured products and discard those with low quality without affecting the supply chain and sales.

- **Emotional Mapping:** Sentiment analysis and emotion detection are one of the most exciting features of NLP. Early NLP systems allowed organisations to collect speech-to-text communication without accurately determining its full meaning. Today, NLP approaches can sort and understand the nuances and emotions in human voices and text, giving organisations unparalleled insight. Learning customer expectations and operators’ viewpoints is a very important element in manufacturing. NLP technologies permit to identify emotions and the polarity of the opinions of customers and operators and provide actions to improve products and different processes. For example, knowing the expectations of customers is key to building a longer relationship and creating engagement with them.

- **Operation Optimisation:** Furthermore, NLP technologies can be employed to trace the performance of equipment and improve the interaction with machines. This simplifies the operation of complex systems and can enable Human Machine Interaction where the operator and the machine collaborate in order to optimise processes.

Therefore, by leveraging NLP technologies, both the decision makers and operators can improve the collaboration between humans and machines within the manufacturing sector and increase the knowledge about their systems and processes.

STAR project aims to enable the deployment of secure, safe, reliable and trusted human centric AI systems in manufacturing environments. Many of these AI systems require interaction with humans and machines and can often benefit from NLP techniques. For example, Speech-to-Text and Text-to-Speech capabilities can enable multimodal interaction with the system, or sentiment analysis can evaluate the polarity of the messages the system receives and adapt to the user’s mood. These user-centric ideas are within the NLP activities of STAR.
The United Nations defines Capacity Building as “the process of developing and strengthening the skills, instincts, abilities, processes and resources that organisations and communities need to survive, adapt, and thrive in a fast-changing world”. In the manufacturing context, this concept refers to the idea that individuals should be able to identify their strengths and abilities, and find ways to enhance them, while adapting to the ever-evolving technological landscape.

This need manifests itself at various stages of a professional career. For example, when attempting to enter the sector, it is important to have a clear idea of the requirements to be eligible for a particular role or to understand what skills are required to perform tasks of interest. If the individual has been employed in the sector for an extended period of time, they may wish to progress to a new role or even change their profession completely. Therefore, it is beneficial to be aware of the similarities between current and future occupations. Furthermore, if the individual is seeking a new role, it is also beneficial to understand how to adapt to the new requirements. In our fast-changing world, learning new skills is rather important for workers. It helps them to keep up with the latest technology, making sure they stay good at their jobs and can handle new challenges.

One of the questions that arises when trying to discover these capabilities and needs is where to find reliable and relevant data. It is true that there is an infinite amount of information spread all over the Internet. It is also true that the presence of AI-powered conversational agents and chatbots can provide information of interest. But there is always the question of whether this information is relevant, based on reliable data or whether it is a hallucination of the AI model.

For this reason, it is key that AI tools that help in capacity building are based on reliable and relevant data sources. One of them is the well-known occupational information database O*NET (https://www.onetcenter.org/database.html), which has been developed under the sponsorship of the US Department of Labor/Employment and Training Administration and is the primary source of occupational information in the United States. With more than 20 years since its first version, with periodic updates and a well-defined data collection program that includes both workers and employers, and with the certainty that being used as a source for training programs and labor studies in several countries, it provides high-quality and credible occupational information.

This database includes a considerable amount of occupational information, including occupations and their alternative names, and entities related to skills, tech skills, knowledge, and tasks. This is the main and most relevant information included in O*NET, and also the most used by researchers and application developers.

ESCO (https://esco.ec.europa.eu/), European Skills, Competences, Qualifications and Occupations, is the version adapted to the European labour market, also including information on education and training. Among many other functions, ESCO is being used by public employment services in different European countries to design multilingual job profiles or for matchmaking between work experience and skills. The European Commission updates ESCO regularly, including mappings to international classifications such as ISCO-08 (https://www.ilo.org/public/english/bureau/stat/isco/isco08/).

At STAR Project, as part of the research tasks on Natural Language Processing, conversational agents for capacity building and the development of the worker training portal, several applications have been developed based on these databases. Specifically, both the STAR virtual interviewer and the occupational information chatbot are based on O*NET as the main data source, and are extended through different AI and NLP techniques so that the user can interact with this reference information in a natural way.

In conclusion, these databases are being increasingly used to develop AI tools for the manufacturing industry. Tools that can help in the knowledge of individual skills and allow workers to train and adapt to changes. In fact, they are of great importance so that the data and suggestions offered by AI based software tools are aligned with reliable and reference sources.

The Occupational Information Databases are essential sources of information for workers, employers, and policymakers. They help in identifying and understanding the skills and competencies required for various occupations, and in matching workers with suitable job opportunities. This is crucial in a fast-changing world where new technologies and industries are constantly emerging.
Extending Factory Digital Twins through Human Characterisation in Asset

By: Vincenzo Cutrona, Elias Montini, Paolo Pedrazzoli, Niko Bonomi / SUPSI-SPS Lab, Giacomo Delinavelli / Arthur’s Legal Strategies & Systems, Tamás Ruppert / University of Pannonia

Owing to the research endeavours undertaken in the STAR project, the team at SUPSI has augmented traditional factory digital twins by incorporating the dimension of human characterization within the Asset Administration Shell (AAS). This enhancement establishes a foundational framework for human-centric control and management systems, as evidenced by the application of an augmented AAS prototype in two illustrative use cases.

Primarily designed for facilitating information interchange across enterprises, the AAS enables the establishment of common semantic frameworks. Nonetheless, semantic definition remains a developing area and is operationalised through the creation of specialised sub-models that delineate the nature and format of information exchange. In this scholarly work, the SUPSI Team introduced an array of sub-models aimed at offering references and guidelines for leveraging AAS in the transition to Industry 5.0 and human-in-the-loop frameworks, resulting in a Human-Asset Administration Shell (H-AAS). These sub-models emanate from a meta-model explicitly designed for the assembly of Human Digital Twins (HDTs). This meta-model outlines the principal classes that require modelling to construct a comprehensive HDT, encompassing factors such as worker attributes, emotional and medical states, psychophysical conditions, and geospatial parameters.

In the context of Industry 5.0, a meticulous digital portrayal of human operators serves as the foundation for data-driven decision-making aimed at enhancing the well-being and resilience of operators. Consequently, the AAS has been expanded to encompass dedicated digital models, replete with a property set designed to articulate the characteristics of human operators and their interactions with their immediate industrial environment.

Two exemplar use cases were constructed within a lab-scale manufacturing system framework. In these cases, equipment and devices were modelled in conformity with AAS standards and interfaced using MQTT protocols, thereby integrating seamlessly with our proposed human-centric AAS extension. Operators were furnished with wearable sensor technology and an informational dashboard, thereby receiving real-time feedback and notifications related to the manufacturing milieu.

Acknowledging the invaluable contributions of Arthur’s Legal Strategies & Systems, the work also discussed ethical and regulatory considerations as an integral part of this augmentation process. The discourse underscores that, although the augmented AAS has reached a level of maturity sufficient for the integration of human operators, existing regulations have yet to catch up with these technological advancements.

Get more info in the research paper available at this link.
Part III:

Industrial Systems Safety
In the STAR project, we try to bring the safety zone detection concept to the dynamic environment. In the current scenario of the SmartFactoryKL testbed, we utilise the robot’s LIDAR sensor for its safety through movement. In the as-is scenario, there is no prediction anticipated. Hence, it does not reach the dynamic layout changes. When the obstacle is close enough to the head side of the Robot, it will stop further movement.

The STAR project brings the detection and prediction phase in the safety zone detection. To achieve this, three use cases were driven in the STAR project:

2. Robot reconfiguration is based on the dynamic layout.
3. Dynamic path planning using both first and second use cases.

In the figure below you can see the diagram of these three use cases insertion into the as-is scenario.

The first use case plans to detect human activities and predict their next actions. For this matter, DFKI initiates typical workflows as the workers’ scenarios, happening during normal daily work. The behaviors of more than 10 participants were recorded, who were supposed to follow the same or similar flows in the free order. The recordings were made using wrist sensors its data are then analyzed in detail to detect the activities the humans are currently performing.

The second use case is to dynamically update the navigation route of the mobile robot, by considering human and/or other (non-)moving objects in the environment. This use case will also enable easier reconfiguration of the robot in case the layout of the environment (including the production stations) changes. The layout is actively monitored by the cameras installed stationary, and humans, as well as the objects in the layout, are detected. In case of any change, the new coordinates of the stations, where the robot should navigate, are updated.

For the third use case, these two use cases are going to be combined to have a safe environment for the workers and the hardware equipment. The newly received coordinates of the stations will be used to set the robot’s destinations. The speed of the robot and the objects in the layout will also be considered to create a collision-free navigation path for the robot. The human’s current and next activity is also one of the important aspects to take into account during the decision. In consequence, the workers’ behavior prediction from the first use case is taken into account here. Here, the STAR project will bring an indispensable functionality and aspect to the industrial environments for the safety zone detection and undemanding reconfiguration of the layout.

To improve human robot cohabitation, a first prototype of a software to detect dynamically security or empty zones throughout the infrastructure using a global situation assessment was developed. For this, we implement AI based algorithms to analyze the scene using the global point of view of a static camera network already deployed in the factory.

Video analytics allows to exploit automatically the video streams in real time to detect anomalies and to raise immediately an alarm. To this end, the algorithms detect, track and localise elements of interest (such as people, robot and new object occupying the scene) over the time. This information will be integrated in another engine to alert the robots of the presence of any obstacles in the surrounding area, in such a way that the system will decide whether a new robot ‘path should be calculated to reach the docking station or to stop completely to avoid any collision.

More in depth, the Safety Zones Detection System exploits video footprints as input and will deliver the spatial heatmaps as results of the analytics. This process combines 2 main components as presented in Figure 1:

- The elements extraction module;
- The 3D object localisation.

Particularly the elements extractor engine merges two deep learning algorithms, either for the skeleton reconstruction in order to follow the human gesture and pose and the other for the detection and classification of non-static object in the scene, with a background subtraction module.

This latter method allows to assess the difference between the background model and the current image in order to infer moving elements in the scene under observation. The 3D object localisation takes as input the results of the elements extractor in order to localise them using Euclidian reference system. To archive this task a calibration software was developed in order to make a correspondence between the camera pixels and the physical word.
Improving the intelligence of the robot with human activity tracking

By: Volkan Gezer / DFKI

With STAR project, we would like the AMR to reach objects and humans, differently. We use different AI techniques to make this possible. First thing is to record daily human behaviour to have an idea of different activities. We iterate the recording by including more people who are performing same activities (in the same and different order) and to collect more data to increase the accuracy. Based on the collected data, we will feed the robot with the possible next activities of the humans, online. The robot will then react to humans differently than an inanimate object. For example, if a human is likely to intersect with the route of the robot in the next two seconds, robot may take another route. Later, this use case will also be supported and merged with the cameras that are feeding live images from the environment.

At SmartFactory-KL laboratory, the autonomous mobile robot (AMR) assists worker by transporting the product between the production islands. Currently, if there is an obstacle in the direction of moving, the robot stops and waits for a specific amount of time before moving further. However, if the object is still there, the object must be removed and AMR must be resumed, manually. The as-is scenario can be seen below:

As modern machine learning algorithms grow in complexity and sophistication so does their need for larger sets of training data. In practice these sets might be smaller than required, be of bad quality or misrepresent the actual domain of the real-life problem that is being modelled. Simulation is a useful tool for counteracting those issues, as it can not only augment the training input and speed up the training process of the algorithm but can also help with its subsequent validation and robustification by producing original samples and scenarios intended to enhance the algorithm’s predictability and trustworthiness. Our aim in the STAR project is to provide a simulated reality module that can work as an add-on for predictive methods, while also interacting with the human operator in order to produce verifiably realistic scenarios that will augment the automated decision-making process.

Simulated reality can be applied in a variety of manufacturing scenarios, such as visual quality inspection and autonomous robotic control. In a supervised setting such as defect detection it takes the form of synthetic data generation. In such a use-case data augmentation is often necessary due to the class skew caused by the rarity of defective artifacts. Underrepresented classes can be supplemented with synthetic images constructed from the existing data. Recent works have achieved good prediction rates under skewed training sets using state-of-the-art methods such as Convolutional Variational Autoencoders and Generative Adversarial Networks. A step beyond is the synthesis of novel defects, usually based on prior knowledge from different products. Neural Style Transfer has been used successfully to fuse defect snippets with non-defective images. Such images can be used to test and also enhance the algorithm’s predictive capacity.

In our work in STAR, we are tackling such a characteristic supervised defect detection use case, where the dataset is small and heavily imbalanced with images of defects being rare and hard to obtain. It was therefore important when choosing a suitable data augmentation method to take into account its ability to handle small datasets, something which is not the case when training large autoencoders or GANs from scratch. We are currently examining and evaluating different state of the art approaches that address this issue, such as trying to selectively utilise complex open-source GANs trained or large datasets such as ImageNet. Of course, lack of data is not the only challenge we have come across; high image fidelity is also an important requirement as differences between defects and non-defects can be minuscule. If this requirement is not met, the use of synthetic data can backfire and confuse the classifier further. To overcome this issue, we chose to purposefully fuse synthetic with real data and apply some filtering and post-processing which has led to interesting results. Our next steps are to improve the quality and usefulness of the synthetic data even further and to produce novel examples that will help assess and increase the robustness of the underlying AI models.

Synthetic Data for Better and Safer AI Models

By: Spyros Theodoropoulos, Dimitris Dardanis, Georgios Makridis, and Dimosthenis Kyriazis /

UNIVERSITY OF PIRAEUS RESEARCH CENTER
Machine Learning for Robots Fleet Optimisation
By: Bertrand Duqueroie / THALES

The STAR project will bring new capabilities for dynamic and adaptive Automatic Mobile Robots Fleet management. This will allow the use of such kind of robots in changing environment, without the need of a time-consuming configuration.

Modern factories rely more and more on mobile robots to perform logistic tasks. They are efficient means to move goods from one place to another within the factory. However, they currently require a specific configuration to be used safely within an environment with human workers. Before installing a Robot Fleet, a digital map of the environment needs to be created. Possible paths for the robots are designed once for all. This kind of solution is adapted when the factory layout and working processes are fixed.

STAR is developing new technologies relying on Artificial Intelligence and simulation to bring the adaptive capabilities needed to use Robot Fleet in wider types of factory environments. Our solution consists in:

1. Creating an updated digital view of the environment, thanks to low cost cameras deployed in the factory and advanced Machine Learning to analyses the situation.
2. Anticipating human movements within the factory, thanks to Machine Learning trained on huge set of factory data, created by simulation.
3. Optimising "on the fly" Robot Fleet commands to adapt to the current layout of the factory and human workers behaviors, thanks to Machine Learning trained by trials and errors, within simulation (Reinforcement Learning).

To keep the cost of our solution low, we rely on a few standard cameras deployed in the factory instead of adding expensive sensors embedded on the robots. Moreover, the occlusion that face embedded sensors will not limit our global situation awareness. Having a clear picture of where are the obstacles (that could be temporary be left on possible paths), and analyzing human presence is mandatory to optimise efficiently robot fleet behavior.

The anticipation of human presence within the factory is a challenge on its own. Machine Learning is key here to extrapolate the current situation in a near future. Like always in Machine Learning, training is crucial. Thanks to our simulation capabilities, we can simulated a huge variety of factory layouts and working processes that will encompass the situations encountered, when our solution will be deployed in a real factories.

Finally yet importantly, we compute optimised Robot Fleet commands, sending the more suitable robot with the most efficient and safest path. Constraints on the fleet can be hard, because of the dynamic environment and because of the need to avoid as much as possible interfering with human movements. Moreover, optimised commands must be very fast to compute, to adapt to the ever-evolving situation within the factory. Once again, Machine Learning, and in particular Reinforcement Learning, is the solution we develop. This kind of solutions as be made famous thanks to impressive results where AI beats professional human players at different games such as Chess, Go, or StarCraft2. Here the AI trains itself in an interactive and fast simulation through a very large number of runs. Thanks to this huge training, it will be able to cope with the large diversity of real situations.
Testing Mobile Robots in the SmartFactory KL e.V. Plant
By: Hooman Tavakoli Ghinan and Volkan Gezer / DFKI

The SmartFactory KL e.V. technology initiative was established in 2005 in Kaiserslautern and now consists of more than 50 organisations from different fields of research. The initiative aims to bring together industrial and research partners within a common network in order to implement joint industry 4.0 projects for the factories.

Being a one-of-a-kind manufacturer-independent demonstrator and research platform, it aims to evaluate and further develop the innovative information and communication technologies in a realistic, industrial production environment.

The SmartFactory KL e.V. focuses on a manufacturer-independent standard, which enables it to have a flexible system expansion by utilising RFID tag standard, an OPC-UA communication platform and standardisation of the hardware, such as different modules with specific mechanical functions, thus facilitating the use of IT systems in the production process.

In SmartFactory KL e.V., the demonstration consists of two main parts.

The first is the Production Level 4. This demonstrator aims to address autonomous production. The human in the production line as a decision-maker is the focus of the Production Level 4. Production Level 4 integrates people and IT as the autonomous elements. The cooperation of these two sides of autonomy enables industries to maximise agility in the production line.

The second, Industry 4.0 demonstrator, aims to provide human-centric support using AI in the manufacturing system. This demonstrator consists of different modules such as storage, transport robot, and supply infrastructure and is used to improve the production line in the areas such as anomaly detection, condition monitoring and predictive maintenance among many others.

STAR project will use the Industry 4.0 demonstrator of the SmartFactoryKL to realise its use cases at DFKI. The mobile robot shown in the figure above is already working autonomously to perform repetitive tasks, such as bringing the products between other modules (stations). It can also stop when it detects obstacles. What we want to achieve in the STAR use cases, is to ease the robot’s reconfiguration, when the coordinates of the modules change. Moreover, we want to increase the safety features of the robot with the help of fixed cameras. Instead of stopping the robot in case of a (possible) obstacle, we want to include reinforcement learning (RL) methods to re-route the robot, thus keeping the production rate high.

First, we plan to record several scenarios in different conditions (e.g., light). Then, based on the collected data, we will train the AI models to automatically detect the coordinates of the modules. Later, by recording some human activity, we predict human’s next activity to estimate the collision possibilities with the robot. Finally, combining these two algorithms, we will control the robot navigation to reach its path without causing any downtimes.

The solutions developed by the STAR project will add new novel functionalities to the demonstrator, which will be also introduced to the SmartFactoryKL lab’s collaboration partners.
Situation awareness for safe robot-human cohabitation in production lines

By: Andreina Chietera and Jean-Emmanuel Haugeard / THALES

One of the main goals of STAR is to ensure the optimisation of a production line to increase the efficiency of the manufacturing process. We start from the assumption that efficiency and safety go hand in hand in a complex environment such as the production lines, in which operators, robots and automatic systems share dynamically the same physical workspace.

The aim of this module is to take advantage of modern computer vision approaches in order to recognise postures and motion of workers, locate them as well as the items occupying the environment. The main output will be an “average spatial heatmap” representing a probabilistic occupancy of the production lines based on fixed RGB cameras deployed in the factory. The purpose of this module is to feed a “planifactor” indicating dynamically which areas should be avoided by robots’ fleet operating in the production lines.

The solution we imagine is conceived by merging dynamically which areas should be avoided by the robots’ fleet operating in the production lines.

and estimation of human-robot distances using the geometric calibration of fixed RGB cameras
• Heterogeneous and homogeneous multi-sensor fusion merging video analytics results coming from cameras dispatched in the production lines including other localisation sensor data.

More specifically, the skeleton detection algorithms allow to track human poses by detecting and estimating the position of the characteristic points defining human postures. These points are parts of the human body (feet, knees, shoulders, neck, nose, eyes). The approach based on a neural network called “Open-Pose” creates heat maps for joint extraction and extracts affinity fields considering all the detected joints in order to infer the link between them and, consequently, allow the detection of human limbs.

Once humans or other items are detected from video footprints, they are located in the infrastructure. This absolute positioning of the elements of the scene requires the camera to be calibrated, in order to associate each pixel of the image space with absolute 3D coordinates system. Once an element is detected, it is projected on the ground, taking into account reference measurements (i.e. height of the body, robot dimension). The projection on the ground allows to estimate the actual 3D position and then the distances between any other elements of the images.

The RGB cameras play a fundamental role and have the advantage to be cheap, extremely robust, and long-term stable.

We think that this AI-based technology will open the door to future optimisation of collaborative human-robot approach to safely share workspaces. The real-time evaluation of the occupancy level of the workspace plays an important role to define the paths robots could cross safely.

In the ever-evolving landscape of robotics, innovation continually pushes the boundaries of how we interact with machines. One such transformative advancement is gesture-based control (GBC), a technology that promises intuitive and seamless communication between humans and robots, revolutionising industries across the board.

GBC allows users to command and interact with robots using natural body movements or hand gestures, eliminating the need for complex interfaces or physical controllers. This technology harnesses the innate dexterity and expressiveness of human gestures, making robot control more intuitive and accessible.

One of the significant advantages of GBC lies in its user-friendliness. The learning curve is minimal since gestures mimic our everyday movements, making it easier for operators to adapt and control robots effortlessly. This simplicity also reduces training time, allowing operators to quickly become proficient in robot manipulation. Moreover, GBC enhances safety in various industries. In scenarios where physical interaction with machinery poses risks, such as manufacturing or healthcare, operators can control robots from a safe distance using gestures. This capability not only ensures worker safety but also optimises operational efficiency.

The versatility of GBC extends across multiple sectors. In manufacturing, it streamlines production processes by enabling precise control and manipulation of machinery. In healthcare, surgeons can control surgical robots with greater precision during complex procedures, minimising human error.

As part of our STAR research, we have integrated and trained machine learning and AI algorithms to enhance gesture recognition, making interactions even more productive and precise. At the present time, a limited class of gestures has been successfully deployed in our factory setup where robots are controlled based on recognised gestures by using HoloLens. Integrating BCD within STAR framework results in improvement safety and human-robot interaction, overall making it a more reliable and safer factory environment. However, main challenges arise from limited ability to process video images at rates which enable safe real-time control.

Ultimately, the ability to merge vision-based approaches and control represents a relevant development in how humans engage with robots in a shared workspace. Its intuitive nature, safety benefits, and diverse applications position it as a transformative technology, poised to redefine the landscape of robotics across industries, opening doors to a more efficient, safer, and interconnected future.

The Power of Gesture-Based Control in Factory Setup

By: Fatos Gashi, Hooman Tavakoli and Nazanin Mashhaditafreshi / DFKI

In the ever-evolving landscape of robotics, innovation continually pushes the boundaries of how we interact with machines. One such transformative advancement is gesture-based control (GBC), a technology that promises intuitive and seamless communication between humans and robots, revolutionising industries across the board.

GBC allows users to command and interact with robots using natural body movements or hand gestures, eliminating the need for complex interfaces or physical controllers. This technology harnesses the innate dexterity and expressiveness of human gestures, making robot control more intuitive and accessible.

One of the significant advantages of GBC lies in its user-friendliness. The learning curve is minimal since gestures mimic our everyday movements, making it easier for operators to adapt and control robots effortlessly. This simplicity also reduces training time, allowing operators to quickly become proficient in robot manipulation. Moreover, GBC enhances safety in various industries. In scenarios where physical interaction with machinery poses risks, such as manufacturing or healthcare, operators can control robots from a safe distance using gestures. This capability not only ensures worker safety but also optimises operational efficiency.

The versatility of GBC extends across multiple sectors. In manufacturing, it streamlines production processes by enabling precise control and manipulation of machinery. In healthcare, surgeons can control surgical robots with greater precision during complex procedures, minimising human error.

As part of our STAR research, we have integrated and trained machine learning and AI algorithms to enhance gesture recognition, making interactions even more productive and precise. At the present time, a limited class of gestures has been successfully deployed in our factory setup where robots are controlled based on recognised gestures by using HoloLens. Integrating BCD within STAR framework results in improvement safety and human-robot interaction, overall making it a more reliable and safer factory environment. However, main challenges arise from limited ability to process video images at rates which enable safe real-time control.

Ultimately, the ability to merge vision-based approaches and control represents a relevant development in how humans engage with robots in a shared workspace. Its intuitive nature, safety benefits, and diverse applications position it as a transformative technology, poised to redefine the landscape of robotics across industries, opening doors to a more efficient, safer, and interconnected future.
Part IV: Explainable and Trustworthy AI
The fourth Industrial revolution (Industry 4.0) has resulted in the automation of many manufacturing processes. Artificial intelligence (AI) models offer astonishing performances in various industrial use cases such as predictive quality management, effective human robot collaboration and agile production.

However, such high accuracy comes at the cost of low interpretability. As interpretability or explainability, one refers to the notion of explaining and expressing, in an intuitive manner, an AI model. In real world applications, AI solutions need to operate as high-performance models which contain huge amount (up to thousands) of hyper parameters which induce extreme internal complexity by using non-linear transformations. To that end, AI models tend to operate as “black-boxes” with a low level of clarity of their inner processes, especially to non-IT experts and other stakeholders, thus generating an issue of trust. The field of Explainable Artificial Intelligence (XAI) has been touted as a way to enhance transparency of Machine Learning (ML) models and increase human cognition.

Stakeholders demand explainability for several reasons. Data scientists use XAI to debug their models and identify why it performs poorly on certain inputs as well as to engineer new features, drop redundant ones and improve model performance. Other individuals use it to monitor their models and be alerted when significant drifts relative to the training distributions happen. Explanations for end users are meant to increase model transparency and comply with various regulations. The importance of explainability as a concept has been reflected in legal and ethical guidelines for data and ML. Specifically, in cases of automated decision-making, the European General Data Protection Regulation (GDPR) require that data subjects have access to meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

The STAR ICT-38-2020 project aims at resolving and overcoming the above issues. Several explainability algorithms are researched and implemented to boost transparency and openness of deployed models in manufacturing processes. Algorithms that identify the most dominant features of deep learning classification mechanisms are among the tools that are used to boost interpretability of the system to stakeholders such as ML engineers and end users (domain experts etc.). Moreover, XAI algorithms that explain the interactions between human and robots will be implemented as well as techniques to identify cyber-attacks such as drifts in the training distribution. In the STAR ecosystem, explainability algorithms are also combined with other powerful concepts such as visual analytics and simulated reality techniques that aim to minimise the risks of physical damage caused by potential agent errors or malfunctions.

As the manufacturing industry continues to integrate AI and ML to improve operations, it is becoming increasingly important to design human-centered XAI for manufacturing applications. The traditional XAI approach prioritises technical aspects and transparency, but the human-centered approach focuses on what, when, and how to explain AI decisions to human end-users, by iteratively involving them in the development process.

The human-centered approach involves techniques such as interviews, hypothetical scenarios, focus groups, and questionnaires to uncover what information is understandable and useful to humans. By involving users in the development process, we can design AI systems that prioritise transparency, interpretability, and user-centeredness. This approach prioritises the user’s needs and preferences and ensures that the AI system is transparent and interpretable to them. Additionally, the use of scenarios early in the system development process to identify user needs for explanations, which can then serve as a basis for further development of explanations. By using this approach, manufacturers can ensure that AI systems are designed to meet the needs of human workers, resulting in safer, more efficient, and more effective operations.

Another area where human-centered design XAI is important in manufacturing is collaborative robotics or cobots. Cobots are designed to work alongside humans, often in close proximity. Human-centered XAI design considers not only the technical aspects of cobot design but also the social and cultural factors that influence human-robot interactions. This approach helps ensure that cobots are used in a way that is safe, efficient, and effective for humans.

The STAR ICT-38-2020 project aims at leveraging human-centered approaches to XAI, focusing on uncovering what, when, and how to explain AI decisions to human end-users, by involving them in the development process. By prioritising transparency, interpretability, and user-centeredness, human-centered XAI can help to improve safety, reduce the risk of errors and accidents, and increase the efficiency of operations in the manufacturing industry.
We should avoid bias in decisions. You recognise that? Yes, it's part of human decision-making recommended practices. But in the last several years, such questions are increasingly posed in Artificial Intelligence – driven decision making. From the "Labeled faces in the wild" dataset, where it took a painstaking 12 years to recognise how unrepresentative the human faces image data sample had been, to the "word embeddings" debate that considers gender or racial bias already present in textual data, issues pertaining to bias in AI-driven decisions are now at the heart of relevant research. Much as "fake news" may be spread by social-network repetition and therefore amplification, bias amplifications in Machine Learning outcomes can be very profound too. One of the most recognisable examples is the COMPAS system case. Its formal adoption led to introducing (automated) bias in the criminal justice system, only to be taken off the system when its dramatic effects became all too clear.

Relevant lessons learned have moved from research and scholarly debates to practitioners on how to handle bias in AI systems and specifically in:  
- Data creation / preparation  
- How a problem is posed  
- How data is analysed  
- How AI-outcomes are validated and tested

But why bias in data-driven decisions only became a major issue in recent years? After all, bias handling in early statistical and control system research and application practice was recognised as an issue to deal with many decades ago. Take any textbook in statistical learning and estimation theory and you immediately recognise techniques for "unbiased estimators" of first and second order statistics. The same applies to control theory textbooks, where unbiased estimators have been put forward when making inferences based on available observations, aiming to drive control actions.

What is the difference now? Earlier literature focused on systems approaches but had a distinctively non-human direction: a system's view would mostly consider performance and operational targets. In contrast, AI is now increasingly making impact on human lives. It's a measure of the great success of the real-world applicability of AI outcomes that this huge leap forward is now an evolving reality. But are we taking a leap-forward to a not properly controlled future? Should there be such a "properly-controlled" future anyway? So much has evolved since the "Three Laws of Robotics" have gained eminence, and yet fundamental problems have not been solved – in fact they are merely starting to be posed. For example, the European Commission, following the recommendations of High-Level Expert Group proposed a framework for Trustworthy AI based on three pillars: ethical, robust, and lawful AI and proposed an assessment list of criteria for Trustworthy AI:

- human agency and oversight  
- technical robustness and safety  
- privacy and data governance  
- transparency  
- diversity, non-discrimination and fairness  
- environmental and societal well-being and accountability

Bias is a key consideration in many of the above criteria. Specifically, from technical robustness and safety, to diversity, non-discrimination and fairness, bias reduction, avoidance or elimination are key aspects to consider.

For us in the STAR project, these are highly important criteria to consider. They are part of the criteria identified in our Human-Centric Design Methodology, are included in our evaluation methodology, and are being considered in the way the human involvement in the AI outcomes is being designed, for example through the use of Active Learning.

But let us take a step back for a moment. Petty much as one cannot eliminate bias in human thinking, wouldn't be appropriate to refocus part of the debate in AI not in the direction of eliminating bias but properly managing bias? In some cases, it may even make sense to seek to strengthen the bias in order to effectively deliver outcomes more consistent with our aims in domain-specific contexts. To do so, we need to recognise bias and work on bias management. The contribution of some of the existing and established sciences towards AI can be important on how bias is managed. Knowledge from medical, production systems, or social sciences-based domains can all be incorporated to bias AI decisions towards domain-specific knowledge and constraints.

Ultimately, a key question is how to "bias" AI to produce outcomes more consistent not just with operational and performance targets, but consistent also with human values. Societies have long accepted humans making hard decisions: the humans making such decisions are ultimately responsible for them. We are now moving towards requirements for human agency and oversight on AI-driven decisions. These are still early days in the ethical debates regarding AI-based outcomes. When the time comes for societies to accept AI decisions on a par level with human decisions, then that should imply that we have properly identified ways of managing the bias of AI solutions in the direction of outcomes consistent with human values. We are still far from meeting such an aim. Still the STAR project is aware of such challenges and contributes towards delivering Human-centric AI. That's why our technical approaches but also our evaluation methodology, both take on board such aspects, aiming to deliver Human-centric AI in production environments and we plan more in that direction.
Industry 4.0 is the automation of traditional manufacturing and related industries, using modern technologies and controlling the industrial value chain. The increasing digitalisation of manufacturing has accelerated the flow of information. Technologies such as Cyber-Physical Systems (CPS), Industrial Internet of Things (IIoT), and Artificial Intelligence (AI) add value to Industry 4.0 value chains. In this way, products and means of production are networked and can "communicate," enabling new ways of production, value creation, and optimisation in real-time.

One of the project goals of STAR is to research and integrate leading-edge AI technologies such as active learning systems, explainable AI, human-centric digital twins, and much more, to allow the safe deployment of sophisticated AI systems in production lines. The Jožef Stefan Institute main competences in the STAR project are in the areas of data analytics and machine learning. We develop methodologies and approaches for active learning. Active learning (AL) is usually the natural approach to provide human-in-the-loop (model requiring human interaction) functionalities in advanced AI systems. Typically, AL attempts to improve learners’ performance by asking questions to an expert to obtain labels for data instances. Since users are often reluctant to provide information and feedback, AL is used to identify a set of data instances on which the provided users’ input conveys the most valuable information to the system. In the decision-making process in manufacturing, AL can also be implemented in recommender systems. In such cases, it tackles obtaining high-quality data that better represents the user’s preferences and improves the recommendation quality. The ultimate goal is to acquire additional feedback that enables the system to generate better recommendations. Collecting feedback from forecast explanations can be realised with a framework of three components: a forecasting engine, an explanation engine, and a feedback loop to learn from the users. We extend this approach to collect feedback from forecasts, forecast explanations, and decision-making options we recommend to the users.

We have developed a system that can acquire and encapsulate complex knowledge. The system is based on semantic technologies, considering ontology concepts that are generic and ported to multiple use cases. It integrates demand forecasting models, explainable AI (XAI), a decision-making recommender system, and a knowledge graph. The components mentioned above are used to develop decision-making workflows displayed through an interactive user interface. Feedback is collected from users regarding forecasts, forecast explanations, and decision-making options shown to the users.

The system requires at least eight components:
- Database - Stores data from manufacturing plants.
- Knowledge Graph - Stores data ingested from a database or external sources and connects it, providing semantic meaning.
- Active Learning module - Aims to select data instances whose labels are expected to be most informative to the system and thus help enhance the AI model’s performance.
- AI model - Aims to solve a specific task relevant to the use case, such case.
- XAI Library - Provides some insight into the AI model’s rationale to produce the output for the input instance considered at the task hand.
- Decision-Making Recommender System - recommends decision-making options to the users.
- Feedback module - collects feedback from the user and persists it into the knowledge graph.
- User interface - provides relevant information to the user through a relevant information medium.

The current work presents a system’s conceptual design to acquire and encapsulate complex knowledge using semantic technologies and AI. The system is demonstrated on a demand forecasting use case in manufacturing. The methodology can be extended to several use cases in manufacturing. The system provides forecasts, forecast explanations, decision-making options, and the capability to provide implicit and explicit feedback. It enables the development of an active learning module that can improve data collection by identifying promising data instances that, when labeled, are expected to be most informative to the system. Our future work in the STAR project will focus on implementing an active learning module and explore recommender systems that learn from data to provide decision-making options to the users.

Stay with us and follow the latest news!
The “little” challenges within the CHALLENGE

By: Mihail Fontul & Gil Oliveira / IBER-Oleff

Iber-Oleff’s (IBER) main challenge within the STAR project is to become agile by good use of AI in the production process. This article is all about the little challenges they face in this journey.

IBER was born 25 years ago, with a small team of 25 people. During IBERs lifetime new projects arrived and thus IBER evolved to what it is today, a 500+ people company acting across several industries always with a special focus on technical plastic parts with kinematics. Although there were improvements in the workflows and processes over the years, the general philosophy and organisation remained the same. Those were the days of mass production where customisation was still a buzz word. Nowadays, the world has followed the mass of mass customisation instead and organisations like IBER must quickly adapt to this trend or they run the risk of getting out of the business. In the STAR project, IBER saw the opportunity to run a first pilot of the new production cells with higher degrees of agility and autonomy.

The first of these ”little“ challenges were related to creating an island of customisation within an organisation that used to work towards mass production. It is like a piece of sand in a huge gear box. The IBER STAR team had to start from scratch a work with company colleagues to start this mind shift without jeopardising the running production. This first challenge brought a good set of lessons learned that training for technical teams and awareness for production teams were much appreciated and had quite an impact. Nevertheless, this is still an ongoing “little“ challenge as the anonymisation of the data is still something IBER is working hard on. The generation of Giga bites of information on real time are being stored in an internal database and must be transformed into valid information to be shared with the STAR partners to be further processed in the AI machine.

In summary, STAR has given IBER the chance to create not only a great pilot project for the use of AI in manufacturing processes, but also to begin a digital transformation journey. The IBER team is excited with what the upcoming months will bring, but also to see the impact of the pilot project in the full extent of the company.

The team was able to understand this little challenge and in a reasonable amount of time span, to adapt the pilot areas to be ready and have the necessary technological upskill to give the mandatory output to enable the project to succeed.

The second “little” challenge was the technological state of the art of the production lines that was in line with the philosophy of mass production. Lack of sanitisation, no real-time monitoring in each work-station and so many other examples, are symbols of the history of the company. To be able to prepare the pilot, changes and adaptation of the work environment were far beyond what was forecasted.

The third and last little challenge is about DATA! IBER never had to share production data with third parties. First because of the severe NDA with clients, secondly because it was never needed! Besides these two main reasons, IBER is a manufacturing site with not so many IT experts and IT expertise, meaning that training for technical teams and awareness for production teams were much appreciated and had quite an impact. Nevertheless, this is still an ongoing “little“ challenge as the anonymisation of the data is still something IBER is working hard on. The generation of Giga bites of information on real time are being stored in an internal database and must be transformed into valid information to be shared with the STAR partners to be further processed in the AI machine.

In this direction, STAR has produced a practical framework for auditing the trustworthiness of AI systems in the form of a “scorecard”. The framework/scorecard serves a dual objective: On the one hand it enables manufacturers and providers of industrial solutions to successfully implement, deploy and operate trusted AI solutions. These complementary assets include the means for assessing, benchmarking and improving the trustworthiness of AI systems and solutions.

Considering the above listed components, the AI trustworthiness auditing process involves the following steps:

- An AI Trustworthiness Evaluation Guide, which provides the means for processing the information of the self-evaluation form. It comprises rules for scoring the trustworthiness of an AI system, based on the information of the self-evaluation form about the system.
- Two main outputs of the AI auditing process, which are produced based on the processing of the self-evaluation questions in-line with the rules and instructions of the evaluation guide.

The auditing framework consists of the following components:

- A Self-Assessment / Self-Evaluation Form for the trustworthiness of AI systems. It consists of a set of questions that relate to different aspects of the trustworthiness of an AI system.

Figure 1: STAR Trustworthiness Auditing Framework Overview
The completion of this step requires that the user of the framework has a good understanding of the AI system and of AI technology in general.

- Scoring of the System’s Trustworthiness: The supplied answers to the various multiple-choice questions are properly analyzed and a trustworthiness scored is computing. The computation is based on the scoring guide of the Auditing framework.

- Provision of Feedback for Improvement: The AI system is aimed at boosting a continuous improvement discipline for the trustworthiness of the AI system. This is based on the scoring guide of the Auditing framework, as well as a plan for potential standardisation of this auditing tool. The following figure illustrates the different aspects that are considered by the STAR trustworthiness auditing framework, which include transparency, explainability, fairness, bias mitigation, robustness, ethical guidelines compliance, documentation, human oversight, data collection, data storage, data management and data quality aspects. As evident from the presented list, several aspects have to do with the trustworthiness of the data that are used for developing, training and executing AI systems.

To deal with these trustworthiness aspects, the framework includes the following multiple-choice questions:

Q1: How does your AI system ensure the transparency of AI models and algorithms?
Q2: How does your AI system ensure the explainability of AI models and algorithms?
Q2: How does your system collect data within the AI system to ensure privacy, security, and integrity?
Q3: How does your AI system ensure the explainability of AI models and algorithms?
Q4: How does your system store data within the AI system to ensure privacy, security, and integrity?
Q5: How does your system manage data within the AI system to ensure privacy, security, and integrity?
Q6: What measures do you implement to ensure the accountability of the AI system’s decisions i.e., to attribute these decisions to specific algorithms or components?
Q7: What measures do you implement to identify and mitigate AI bias situations?
Q8: What measures do you implement to ensure the robustness of the AI system?
Q9: What measures do you implement to ensure the fairness of the AI system?
Q10: What measures do you implement to ensure compliance with ethical standards and guidelines in manufacturing?
Q11: What measures do you implement to ensure the quality of data of the AI system?
Q12: What measures do you implement to ensure the cyber-security of the AI system?
Q13: What measures do you implement for human oversight and intervention when necessary to ensure that AI decisions align with human values and intentions?
Q14: What measures do you implement to provide comprehensive documentation for the AI system?

Based on the answers to the above-listed questions, a trustworthiness score is calculated. In this direction, it is assumed that the more measures an organization takes regarding the trustworthiness of an AI system, the greater the trustworthiness score of the system. STAR provides the scoring guide in-line with this approach, including the upper and lower margins of the trustworthiness score for each question and for the auditing framework (“scorecard”) as a whole. The presented approach offers the following advantages:

(i) It is very simple and easy to understand; and
(ii) It can score different systems automatically based on clear and unambiguous scoring rules. However, it also suffers from problems and potential inaccuracies as it assumes that all measures and questions are of equal importance, while giving extra benefits to measures that can foster different dimensions of trustworthiness. To alleviate these issues, it is possible to configure the scoring process in different ways by assigning some weights to specific measures and/or questions.

This first version of STAR’s trustworthiness auditing framework, which will become soon available in the STAR Market platform. We are in the process of improving the auditing framework based on feedback from STAR project members and other industry stakeholders. We also plan to provide complementary assets that will boost the sustainability and wider use of the framework. Such assets include a mini training tutorial about the different questions of the auditing framework, as well as a plan for potential standardisation of this auditing tool.
Part V: AI Cybersecurity
Halfway to Security and Data Governance for AI Systems in Manufacturing

By: Dimitris Papamartzivanos / UBITECH

One of the main objectives of STAR project is the provision of mechanisms that will ensure the security and data reliability for AI systems in manufacturing. STAR consortium has worked towards achieving this goal, and we are pleased to announce that STAR partners have completed the first version of the AI Security and Data Protection layer of the STAR’s overall architecture.

The figure 1 illustrates the internal architecture of the AI Security and Data Protection layer of STAR, which aims to bridge the gap between the manufacturing plant and the factory security officer by increasing her awareness regarding the cybersecurity posture of the production lines.

Several individual components work in synergy like rolling engine gears to convey evidence from the manufacturing environment to the security officer. The goal is to enable informed decision making for mitigation actions and ensure the timely adaptation of the production procedures so that to ensure business continuity and environment’s safety. These gears are:

- **AI Cyber Defense for Secure and Trusted AI algorithms**: STAR develops AI technologies that secure the operation of the AI systems and algorithms that they comprise. In this direction, the project implements AI Cyber Defense tool that protect and defend AI systems from malicious security attacks. The goal of STAR focuses primarily on defenses against poisoning and evasion attacks against AI-enabled systems. UBITECH has worked for the definition of the architectural design of the AI Cyber Defense tool, as well for the evaluation of state-of-the-art attacks and defenses in the context of the actual STAR pilot environments, leading to the completion of first prototypes system and a scientific publication16.

- **Runtime Monitoring System (RMS)**: Enables a real-time service that collects security-related data from monitored IoT system components or applications and stores them for further processing. Analytics algorithms analyze the collected data to detect abnormal patterns. Additionally, the collected data feed the logic of the Security Policy Manager which reports incidents exceeding “normal” thresholds. The system is capable of deploying different monitoring probes responsible for the data collection and publishing to the monitoring platform. Netcompany-Intrasoft has already delivered a prototype of the RMS.

- **STAR Security Policies Manager (SSPM)**: Is a tool used to enforce the logic on the detection of abnormal events in a manufacturing environment. The tool has been designed by GFT ITALIA Srl and enables the security/IT officers to configure security policies according to specific business and security requirements. The main purpose of the SSPM is to correlate the evidence collected from the RMS and the AI Cyber defense tool and report the detected cyber security incidents to the risk assessment module.

- **Risk Assessment and Mitigation Engine (RAME)**: Complements the SSPM for the visualization of the threats and the corresponding risks. The RAME is based on OLISTIC, UBITECH’s Risk Assessment tool which can support the security officer on getting an overview of the security status of the factory, and more specifically, of the production lines and business processes of interest. Overall, RAME enables the risk management and the identification and visualization of risks through comprehensive and reactive visualization.

- **Distributed Ledger Services for Data Reliability (DLSDR)**: Provides the means for tracking and tracing industrial data for AI algorithms, notably the definitions of the data sources used, the data used to configure STAR AI algorithms and finally the data for persisting their results. To this end, it provides services to the AI algorithms and applications utilising their results. Netcompany-Intrasoft has provided a prototype of the DLSDR module in order to reinforce the reliability and the security of the source data used in the STAR system by recording information (i.e., metadata) about the acquired data to facilitate the detection of abuse and tampering attempts against these data.

The above-mentioned components comprise the Security and Data Protection layer of the STAR. At the time of writing this blog, the first prototypes of the components have been delivered, meaning that halfway has been covered! The STAR consortium partners look forward to the exiting next steps for the delivery of the full-fledged Security and Data Protection layer of the STAR!


Figure 1: Security and Data Protection layer of the STAR
STAR’s Blockchain for Data Provenance and Traceability: Tackling the Challenges of Industrial Data Reliability

By: Angela Maria Despotopoulou, Nikos Kefalakis, John Soldatos, Charalampos Ipektsidis / Netcompany-Intrasoft

Understanding Industrial Data Reliability Challenges

Data reliability is a key prerequisite for the development of industrial applications that leverage Big Data, Machine Learning (ML) and Artificial Intelligence (AI) systems. It is also important for ensuring the trustworthiness of AI applications in industrial environments: Without reliable data, it is impossible to develop trusted AI systems. For instance, training ML algorithms with unreliable data will result in models with poor performance. Data reliability is critical for ensuring that ML systems perform as expected.

- Background noise such as noise pollution or interference (e.g., alarms, extraneous speech) and electrical noise from devices like motors, cooling devices, air conditioning, and power supplies.
- Faulty or inaccurate sensors such as sensing systems with poor precision.
- Dying batteries that compromise a system’s ability to provide reliable measurements.
- Compromised or attacked devices that produce biased or fake data due to adversarial attacks (e.g., data poisoning, data modification, false information injection).
- Compromised analytics algorithms, such as algorithms under poisoning, evasion, and other types of adversarial attacks.

To alleviate data unreliability, industrial organisations need data infrastructures that are cyber-resilient and cannot be tampered with. In this direction, the use of distributed ledger technologies is suggested in several research works. Blockchain infrastructures facilitate decentralised data management, including decentralised data operations and transactions. The advantage of a blockchain infrastructure for reliable data operations include:

- Lack of single points of failure: Blockchain infrastructures operate in a highly distributed fashion and do not rely on a trusted third party for the validation of data transactions. This architectural property makes them much more difficult to be compromised, as they have no single point of failure.
- Tamper-resilience: Blockchains have anti-tampering properties. Data written in a distributed ledger requires a next-to-impossible investment in resources to be changed. This is a foundation for data reliability, as blockchain data cannot be changed by adversarial parties.

- Data transparency and auditability: Transactions that write/store (meta)data on a blockchain are transparent and accessible to all members (peers) of a blockchain network. Hence, they are auditable by other participants to the blockchain network.

- Security: Blockchain infrastructures offer integrity protection mechanisms, including data hashing and cryptographically linking among the various blocks. This boosts their tamper-proof nature and minimises security risks. Also, it is not possible to hack a blockchain by attacking few of its nodes. Blockchains support consensus mechanisms, which require an absolute majority of nodes to agree on changes to the blockchain contents. Therefore, blockchains are resilient against cyber-attacks that could compromise one or more nodes.

Introducing the STAR Blockchain Infrastructure

Motivated by these properties benefits, STAR implements a blockchain infrastructure to boost the reliability of the data that are used by the project’s trusted AI systems. The figure below illustrates how the Blockchain Data Provenance and Traceability service interacts with other non-Blockchain modules of the STAR platform.

The STAR blockchain exhibits a rather complex architecture, the assemblage of which requires the use of several interconnected machines each hosting some of its components, thus formulating a private permissioned Blockchain network. Permissioned blockchain provide much better performance than the popular public blockchain networks (e.g., the Bitcoin network) since they need not employ computationally expensive Proof of Work (PoW) mechanisms. Rather, they leverage Proof of Stake (PoS) mechanisms, which enables permissioned blockchain infrastructures to support thousands of transactions per second. STAR leverages the open source Hyperledger Fabric project from the Linux foundation for the implementation of its blockchain infrastructure.

As illustrated in the figure, an organisation participating in the network in this context is a non-Blockchain module of the STAR architecture (e.g., a Cyber Physical Production System (CPPS)), that gains benefit from recording information on the Blockchain. Everything that interacts with the Blockchain network acquires their organisational identity from their digital certificate and their Membership Service Provider (MSP) definition. Communication of service owners with the Blockchain Network, takes place not directly, but via a multi-level Backend application that exposes various APIs to client applications.

A main value proposition of the STAR blockchain for data provenance when compared to similar block-
chain-based systems like Provchain and ProductChain lies in its ability to record and protect not only raw data and source industrial data, but also AI/ML models and analytics results. This is the reason why, STAR's blockchain for data provenance is suitable for trusted AI systems which is in-line with the overall aim of the STAR project.

STAR's Blockchain for Data Provenance and Traceability: Tackling the Challenges of Industrial Data Reliability

By: Angela Maria Despotopoulou, Nikos Kefalakis, John Soldatos, Charalampos Ipektsidis / Netcompany-Intrasoft

STAR's Blockchain as a Docker Swarm Overlay Network

STAR develops an industrial grade system leveraging a Docker Swarm Overlay Topology. To understand the use of a docker overlay network by STAR, the following three use case sample scenarios involving three organisations as illustrated in the figure can be considered:

- Organisation "one" offers to the platform a data stream addressed to Organisation "two": Metadata on the source of the data are stored on the Blockchain. Organisation "two" that needs at a later point to verify the source of the data stream may refer to the Blockchain.
- Organisation "one" offers to the platform a service leveraging an artificial intelligence algorithm (data processor). Metadata on the processor’s configuration (state) right before its execution are stored on the Blockchain. Organisation "two", another stakeholder of the STAR platform, that needs at a later point to verify which processor and under which conditions performed the data processing may refer to the Blockchain.
- Organisation "one" offers to the platform a service leveraging an artificial intelligence algorithm (data processor). Metadata on the results of the algorithm’s calculations are stored on the Blockchain. Organisation "two", another stakeholder of the STAR platform, that needs at a later point to verify a result they might have come across via a different route may refer to the Blockchain to assert that it has not been tampered.

As evidenced in the figure, Organisations “one” and “two” are hosted by two distinct virtual machines carrying the role of Workers within the Docker Swarm network. For the transition to a production-level deployment, the administrator can replicate those Workers to match the number of stakeholders requiring services from Hyperledger Fabric. The Manager within the Docker Swarm network corresponds to a third virtual machine. There resides Organisation “zero” and a set of Fabric Orderers (accompanied by its Certification Authority). This configuration is not mandatory for the future full-scale model. The latter could employ an odd number of Managers hosting only replicas of the Fabric Orderers and network monitoring/administration tools. In case of failure of the Leading Manager another could take the mantle of overseeing the network processes. For the MVP the Manager hosts also an Organisation just to cut down on resource expenses.

The components that constitute the “Organisation” have already been described: the actual Fabric Peer Node, a Certification Authority, a Command-Line Interface for administration tasks, a CouchDB instance to persist the global state and a Java Spring Boot Application exposing an API with the business logic to the outside world. All shall be deployed in Docker containers. Fabric Channels are, in essence, also materialised through Docker containers. Those can be hosted anywhere but let us assume that they will be hosted also within the Manager machine for clarity. Smart Contracts are also being deployed as Docker containers and associated strictly with Fabric Channels. All Peers and the set of Orderers are then attached to Channels for them to effectively share the global state of the distributed ledger.
Manufacturing could be considered at the historical level the "Digital Twin" of human skills evolution. It reflects our growth and ability to build tools, goods, to overcome challenges. Nevertheless, it is one that very special area of our societies where human mind turned skills in complex tools. Industrial revolution and its disruptive impact over the ages is peered by other probably less observed major changes in societies. Those ones, like accessibility of education, communication and travel means, accelerate, and deliver inputs for new waves of growth, efficiency, and diversity in manufacturing. All those cycles of complexity growth refined the understanding, pre-venting, monitoring and reacting to threats related to safety, security and privacy of humans, machines and processes.

Communication networks and computers are key enablers of last few decades able to bring in the game a big number of process and motion control systems. If until no more than 30 years ago, subjects like Artificial Intelligence, Machine Learning or classic Siemens Simatic controllers. Those approaches have been preferred offering a wide understanding and validation of processes to be executed. Still, analytics technologies offered a new push to Machine Learning and Deep Learning techniques peering them with descriptive technologies, like the ones related to Knowledge Graphs or Constraints Programming, both being able to instrument together problem design hypothesis and operational data collected. The aim of such techniques is to offer the tools usable for validation of execution scenarios before effective deployment in production systems. The STAR project considers not only the instrumental aspects of trustworthiness and explainability of AI but also the ways how the AI solutions do the step from the work of art to systematic industrial use. This means a specific focus on the needs to offer certification mechanism altogether with pre-validated compatibility and reliability application scenarios.

Last but not least, STAR is taking the effort to provide a solid path towards industrialised AI being aligned and contributing to normative efforts at both EU and global level.

Simulation have been perceived as a subject of a rather reduced academic community we have can see now how various industries are building models of action almost completely digitalised. Control and operational data flowing through digital platforms associated to manufacturing processes are no longer managed in closed systems with no connectivity to outside world. Mixed fog-edge -cloud deployments constitute the new reality where functionalities are managed in a DevOps manner. Trust and security are now considered from some different perspectives simultaneously, each of them being relevant for both horizontal functions but also on the vertical applications views.

Up to recent time the level of explainability of decision with many industrial applications stayed within the level of "Human in the loop" control doing enablement of various processing steps (e.g. possible sensitive chemical plant processing or visual inspection of machine tools space), or being based on purely declarative rule based systems as
The advent of next-generation smart connectivity systems has enabled the collection of rich datasets, as part of the manufacturing process, which has emerged as a key enabler for a wide gamut of safety-critical applications ranging from improving predictive maintenance to enhancing cyber-security to enabling remote machine calibration and human-robot collaboration. A key enabler in this direction is the emergence of Artificial Intelligence technologies that can make manufacturing safer and more efficient.

Despite the clear advantages that such unprecedented quantity of data brings forth, it is also subject to inherent data trustworthiness challenges due to factors such as malevolent input and faulty sensors. With constricted demand for manufactured goods and more reliance on remote solutions, preempting machine failure and reinforcing security has never been more important. One of the main hurdles to actively gauging factory assets’ health is on the validation of the received data correctness; especially considering the vast amount of available machine data towards predicting potential failures. Industrial control systems also remain vulnerable to cyber-attacks.

Promising advances in artificial intelligence and machine learning are enhancing the way manufacturers prevent asset failure, block cyber threats, and autonomously calibrate machines. There has been a plethora of proposed solutions, based on the use of traditional machine learning algorithms, towards assessing and sifting faulty data without any assumption on the trustworthiness of their source. However, there are still a number of open issues: how to cope with the presence of strong, colluding adversaries while at the same time efficiently managing this high influx of incoming machine data? Designing such trust anchors is not an easy task: it requires fusing (contradictory) data, originating from untrustworthy sources describing dynamic and uncertain phenomena evolving over space and time.

Particularly with respect to safety and security, system components (managing machine data assets) must be enabled to make and prove statements about their state and actions so that other components can align their actions appropriately and an overall system state can be assessed, and security policies can be evaluated and enforced. This goes substantially beyond simple authorization schemes telling who may access whom but will require understanding of semantics of requests and chains of effects throughout the system and an analysis both statically at design-time and dynamically during runtime. The latter will then even allow to conduct dynamic Risk Assessment (RA) and decide at runtime if an entity is still safe to be used (even if some components are compromised and fail) or needs to be shut down in a failsafe state.

Compounding this issue, the STAR ICT-38-2020 project focuses on leading the design of AI-enabled data verification frameworks with wide applicability in manufacturing environments. It will employ deep learning techniques in the form of neural networks to address the issue of possible data poisoning and evasion attacks in smart manufacturing environments while demonstrating high accuracy and scalability.

With this, we claim that manufacturing supply chains, in general, can withstand even a prolonged siege by a pre-determined attacker with known or unknown capabilities as the system can dynamically adapt to its security and safety state. This is substantially more flexible than traditional security mechanisms that often try to maintain and enforce pre-defined policies using rather static security mechanisms. Even more, STAR’s intelligent multi-layered framework (including Explainable AI, Active Learning, Simulated Reality, Human-Centric Digital Twins for AI) allows a very high degree of automation, something that is definitely required in manufacturing scenarios where the mere number of devices will prohibit human intervention for security management.
Part VI:
AI Ethics and Regulations
AI is not about AI

With any emerging technology, or combination of existing technologies, one tends to focus on the technology itself. However, the technology should not be the focal point as it in itself is not the solution. It is a tool.

AI is no exception. The same goes for high-performance federated and edge computing, and the increasing creation of and access to data.

AI therefore is not about AI. It is about figuring out and helping out addressing challenges – of which each member state, region and society at large have a lot of – and achieving objectives that matter, for people, planet, prosperity, peace and partnership. With this mindset, Europe has great capabilities to lead.

In the first blog of this Series, we discussed the question why: why is the symbiotic, dynamic equation of both functionals and non-functionals (such as security, privacy, maintainability, interoperability, digital sovereignty and accountability) also known as: ‘all-functionals’, one of the main success factors for Industry 5.0 and related value creation.

In the second blog, we focussed on the questions what and where to start: what are the relevant non-functionals, what does success by design mean, and where to start when plotting and mapping the contextually-relevant risk spectra, and appropriate levels of dynamic accountability to cater for those hyperspectral risks.

In this third blog of this Series, we will dive into the question how: how to design the contextually-relevant symbiotic dynamic equation of all-functionals, including some examples on how to methodologically make AI truly useful and work in a trustworthy way.

How to Make It Work, with or without AI?

There is no need or obligation to make AI work. As already mentioned, technology – emerging or otherwise – is one of the many tools we can deploy. This, if and to the extent it makes sense to do so. There is a need to make our vast, vital yet vulnerable supply chains work, and making smart manufacturing, prognostic maintenance and related Industry 5.0 domains work as well. That is the ‘it’ we need to focus on. Where and how to include technology and other dimensions, to make those future-proof supply chains, manufacturing and industry 5.0 truly work?

Dimensions in the Digital Age

In this Digital Age, there is a complex of dimensions to be considered and taken into account, by design, and before, during and after deployments, maintenance and optimisations.

Any start of addressing the question ‘how’ requires a start with identifying a starting point while understanding that there is a lot to take in.

There will be various and numerous dynamics of all sorts and sizes. This, as the already mentioned technological and related developments are expedited by non-digital global occurrences such as pandemics, geopolitical and demographic developments.

There are North Stars to continue to follow, including the recent European Declaration on Digital Rights & Principles19, Europe’s Fit for the Digital Decade Compass20, its Data Strategy21, and Cybersecurity Strategy for the Digital Decade22.

There are objectives, targets, milestones and the like. But there is no end point; no finish. It is and will be required, continuous effort, while catching up and keeping up with the dynamic developments in the various, nine (9) intertwined dimensions, such as structured and visualised below in Figure 1.

It does not matter where one starts. One will generally start in the dimension it is most comfortable with, which is perfectly fine. However, each of these dimensions need to be taken in, covered and – in contextually relevant manner – balanced out. For instance, one can have great emerging technology available, but without the appropriate amount of competences and capabilities of leadership in organisations, and of the persons able to design, deploy, use, appreciate and maintain, the technology will not make it work, and, basically would be useless.

Dimensions of the Digital Age

[Figure 1: Digital Age Dimensions]

21 European strategy for data: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066
Double-Looped Scenario Plotting

To consider risk in order to take technical, organizational and other operational measures by design and thereafter continuously, the scenario plotting methodology of Double-Looped S.I.M. can be used. S.I.M. means: Scenario, Impact, Measures. The double-looping refers to the notion that any measure in itself can be a vulnerability and can even increase risk or create new risk and related detrimental impact. So, every measure deserves its own S.I.M. cycle as well. The Double-Looped S.I.M. can be visualised as set forth below in Figure 2.

Chaos Engineering

The question 'what happens if things go wrong?' is not the one most designers, developers, investors or marketeers wish to ask themselves. Even more, in the AI domain it is expected that incidents will have an even more severe impact than in the digital domains without AI capabilities. These notions also go for other emerging technological capabilities. Good and extensive scenario plotting and mapping are a prerequisite, also from ethical and accountability perspectives.

The appropriate balance between functionalities and their benefits on the one hand, and non-functionals and impact-mitigation and their benefits on the other hand, with appropriate security and other prevention-, risk- and impact-based measures, metrics and measurements in place will need to be found per situation, per context, and meanwhile monitored and challenged continuously. Just keep in mind that there is always another angle. Therefore, designing digital (eco)systems for failure is essential. Design for failure, or chaos engineering helps addressing those various angles and scenarios. It will increase transparency, reduce unpleasant surprises, reduce embarrassing excuses, and most of all increase trust and trustworthiness. Making it work is complex but that is where the true huge potential is.

Co-Creation Cycle: Multi-Disciplinary & Inter-Disciplinary

The Co-Creation Cycle is an aid that identifies the various all-functionals that are relevant in a particular design, development, manufacturing, logistics, monitoring, maintenance or subsequent deployment phases. It helps identifying the various expert stakeholders that should be part of the team in order to find, balance out, arbitrate, document and optimise a symbiosis of the all-functionals that is feasible from technical, operational, economical, ecological, financial, ethical and legal perspectives, as well as otherwise acceptable for all the team members. It further demonstrates that both a multi-disciplinary and inter-disciplinary mindset and skillset is essential to make it work.

The Human-Centric Co-Creation Cycle visualised below in Figure 3 provides for an example where – after identifying the envisioned functionality and related interfaces – non-functionals such as security, safety, authentication, non-personal and personal data control, processing, protection and analytics need to be part of the symbiotic equation by design by design. If the set of desired all-functionals end up being too expensive, too unsustainable or, otherwise, not feasible, the cycle is repeated. It can happen that it needs to be repeated multiple times before – finally – the dynamic symbiotic equation has been established that is deemed – by all stakeholders involved – to be feasible and acceptable for the entire life cycle.

This will be a main success factor in any use case, design, application or deployment, if considered and included

(a) by default by design upstream,
(b) by default at engineering, assembly, implementation, making available midstream, and
Accountability in the Digital Age

With continuously working on the why, what, where and how, as briefly discussed in this and previous blogs, the initial fundamentals of accountability have been laid as well. Accountability is not a mere after-thought, dealt with after something goes wrong. It is an essential requirement, both before one acts as well as during and after. Accountability is about owning and co-owning roles and responsibilities, finding solutions, making things happen, and to helping out if things may go wrong once in a while. Accountability also caters for becoming or being more future-proof while being complaint to relevant ethics, standards and other applicable policy and legal frameworks. In any case, accountability is also not about blaming others. This also as blaming other means giving up the power of change. And change is the only constant, also in this dynamic Digital Age.

In a world were products and services are traded with light speed, trust between partners is a key ingredient. Society, institutional and economical partners has developed norms and standards willing to promote and ensure compatibility and agreed processes. Those activities include terms and their semantics, validation conditions, quality gates, measurements, safety conditions, to mention just a few of them.

With respect to Artificial Intelligence, we could observe that its role along various human related activities exposes some special features. First and obvious one is related to the nature of AI against the governance process which is expected to be executed with ethics and privacy conditions and enable the trust between involved actors.

The AI landscape is complex, and standards and regulations need to carefully structure actions and measurements in order to provide a coherent and applicable framework aiming for trustworthiness.

The European plan for AI regulation took a careful stance and is now represented by the EU Artificial Intelligence Act, the first proposal of this regulation was published in April 2021 (https://artificialintelligenceact.eu).

In simple terms the policy objectives express the objectives to be followed:

- Ensure that AI Systems on EU market are safe and respect EU laws and values
- Create legal certainty to facilitate investment and innovation in AI
- Enhance the governance and enforcement of existing legal requirements
- Facilitate the safe, lawful and ethical development of AI applications
- Harmonised rules for the placing on the market, putting on the service and the use of AI systems in EU
- Prohibitions of certain AI practices
- Specific requirements for high-risk AI systems and obligations for operators of such systems
- Harmonised transparency rules for AI systems intended to interact with natural persons, emotion recognition systems and biometric categorisation systems, and AI systems used to generate or manipulate image, audio or video content.
- Rules on market monitoring and surveillance

The objectives of this regulation are: to ensure that AI systems on the EU market are safe and respect EU laws and values, create legal certainty to facilitate investment and innovation in AI and I facilitate the safe, lawful and ethical development of AI applica-
AI Norms and Standardisation – the road ahead
By: Septimiu Nechifor / SIEMENS

In its current form, the regulation aims to encourage technical regulations such as allocation of responsibilities in the AI value chain between providers, users, brokers, etc.

Yet where is STAR located? The project comes in a sweet spot due to a number of arguments. STAR, by design, considers that AI systems on the EU market are safe and respect the EU laws and values, create legal certainty to facilitate investment and lawful and ethical development of AI applications.

Industrial use of AI, especially under the Industry 4.0 umbrella, aim to provide innovation and productivity increase on areas where humans and machines collaborate such as advanced manufacturing. STAR project use cases and architecture components demonstrate that use of AI can be safe and effective if the AI solutions consider the observability and validation of shopfloor planning and actions.

Lately, the EU AI Act should consider industrial and B2B applications and distinct it from B2C applications and have a close attention to industrial/machine data. This aspect is confirmed by the interest declared by the European Committee for Standardisation (CEN) and the European Committee for Electrotechnical Standardisation (CENELEC) to scrutinise and drive European standardisation on this aspect (see here). At the same time, we should be aware that the AI Act’s declared scope is consistently larger than industrial landscape on its current form. That means that we can only consider use of AI as a high-risk factor when health, safety and security are under scrutiny.

The STAR project admits that a “continuous compliance” with current regulations in place approach is valid and useful. The project’s efforts do not match just a technological promise, but also a plan for evolutionary move towards safe and reliable innovation in manufacturing.

AI Maturity Assessment for Manufacturing
Enterprise: state of play
By: Andrea Ferretto Parodi and Matteo Cardaoli / GFT Italy

Artificial intelligence (AI) is revolutionising how society functions and is opening up new business opportunities for manufacturing companies.

Access to AI tools in organisations is increasingly facilitated by the availability of big data, digitisation, access to efficient and cost-effective computing capabilities, and the development of new algorithms that are increasingly within reach of non-expert users.

Therefore, it is essential for organisations to understand their possibilities to benefit from AI, their knowledge-technical-business gaps, and to put in place the proper corrective measures to fill such gaps. To this end, AI Maturity Assessment methods and tools have been recently developed, as an extension-customisation of Digital Maturity Models (DMMs).

DMMs offer organisations a simple but effective way to measure their capabilities in a given area and contribute to organisational transformation and the development of organisational competencies by identifying a transformation process.

It may be useful to point out that maturity models are different from surveys. A survey responds to the need to provide a picture of the situation to a stakeholder; usually, an institutional one, on the largest and most representative sample of a population of organisations to help the stakeholder understand the sector and decide on actions.

A maturity model, on the contrary, investigates specifically (in a more or less in-depth way) the organisational maturity and the level of technological development of an organisation concerning a discipline. It pursues informative purposes (increase skills, provide experience, raise awareness) and has the ultimate aim of identifying improvement paths for the organisation, starting from a quantification of the current maturity. Maturity model output is generally in a (radar) chart form; this quantitative aspect is fundamental to defining the starting point of a development path. It is also helpful for comparison with other organisations and summary evaluations of an entire sector.

In order to be effective, assessment methods and tools must be sufficiently complete for the subject matter and have inner coherence between questions and in the graduation of the scale of values of the answers, at least tentatively.

Most assessment tools can be found on the web and be filled directly, without assistance. There is also a smaller number of more complex tools that require the presence of a professional both in the investigation phase and in the processing of the information collected. In those cases, the assessment is carried out on-premise. It can be followed by developing a roadmap along the most promising or important lines to improve the organisation’s competitiveness according to its strategic objectives, identifying opportunities and costs. The following steps would be
a gradual integration of the developed solutions into the business processes, up to developing new business models in the most successful cases.

The market offers a wide variety of models assessing the maturity of organisations in terms of digitalisation or Industry 4.0. The number of maturity assessment models related specifically to AI adoption, on the contrary, is still limited and of relatively recent development. We describe here two of the most significant ones.

**VTT AI Maturity Model**

This tool was developed in 2019 by the VTT Technical Research Centre of Finland Ltd. (VTT) and the University of Oulu, based on previous works on the subject. It is a quick self-assessment web tool evaluating the AI maturity level of organisations like SMEs, Large enterprises, the Public sector, and Digital Innovation Hubs. Essentially, it answers the questions "How prepared is your organisation for the use of AI?" and "Where is your organisation, compared to other tool respondents?"

It consists of 12 questions, two for each of the following dimensions: Strategy & Management, Products and Services, Competences and Cooperation, Processes, Data, and Technology.

The VTT model doesn’t require a high level of AI literacy to answer and provides as output a radar chart that visualises the organisation’s maturity at a glance to document the current state of AI and serve as a starting point for subsequent developmental actions.

**AppliedAI Maturity Model**

The appliedAI MM was developed by UnternehmerTUM GmbH in 2021. Unlike VTT and most digital maturity assessment models in use that focus on processes, appliedAI is more focused on how AI is implemented in the organisation, irrespective of the processes it is applied to. It too is a self-assessment web tool devoted to SMEs, Large enterprises, the Public sector, and DIHs, but it requires at least a medium level of AI literacy to respond.

It consists of more than 100 questions, divided into nine areas: **AI Ambition** and **Steering** (referring to AI Vision and Strategy), **Use cases** (concrete, systematic and widespread applications of AI in the organisation, consistent with Vision and Strategy), **organisation** (how the organisation creates the necessary structures for effective collaboration, adoption, control and feedback of AI solutions), **Expertise** (skills enhancement, training, motivation; collaboration with external experts), **Culture** (how the organisation creates an AI-oriented culture), **Technology** (presence in the organisation of infrastructures, technologies, and procedures that support the adoption of AI solutions), **Data** (how the organisation is doing in terms of data - ingestion, cleaning, storing, quality, access...), **Ecosystems** (mainly partnerships), **Execution** (effective and efficient application of AI models to business processes).

By its nature, this tool is aimed at the entities that have already adopted AI solutions to reveal areas to be addressed and provide insightful recommendations of action toward higher levels of AI maturity.

AI’s importance in organisations is rapidly developing and spreading AI maturity assessment tools. It will be up to the organisation to identify appropriate tools and partners to undertake an effective roadmap for AI adoption. For instance, in the Digital Manufacturing Platform cluster, the **Connected Factories “AI for Manufacturing”** new pathway is a tool to position industrial cases in a 5-levels scale of AI-enabled autonomous behaviors, from human to machine-controlled systems. STAR will be present on June 13th afternoon (Brussels) to present and demonstrate the first results of such an assessment (by invitation only) **Connected Factories European workshop on the AI for manufacturing Pathway**.

The **STAR ICT-38-2020** project is one of the H2020 innovation actions aiming at applying highly innovative AI models and solutions to Manufacturing, mainly focusing on technological enablers and pillars for Cyber Security, Safety, and Explainable / Trustworthy AI. STAR provides a set of three industrial pilots: PHILIPS Consumer Lifestyle (Human-Cobot Collaboration for Robust Quality Inspections); IBER-OLEFF plastic components (Human Centred Artificial Intelligence for Agile Manufacturing 4.0), and DFKI SmarFactoryKL (Human Behaviour Prediction and Safe Zone Detection for Routing). Their maturity regarding the STAR advanced AI solutions will be carefully assessed, and gaps identified to maximise the impact of STAR AI solutions.
How to Make Human-Centred AI Work; Not Just Function?

By: Arthur van der Wees / Arthur’s Legal, Strategies & Systems

Will AI bring Social Prosperity? Or Social Disruption?

Intelligent supply chains, rapid innovation production, integrated logistic support, prognostic health monitoring, predictive maintenance and other Industry 5.0 domains have the capabilities to address Societal Challenges and improve productivity, safety, security, sustainability, and other efficiencies.

New concepts, models and processes supported with AI and other digital capabilities are not a nice to have; they are a need to have. For sure it will and should support and augment the workforce, yet it will also challenge and change it, in an evolutionary or revolutionary way. Said otherwise, what one markets as ‘beneficial’ can easily lead to social, and societal unrest and disruption.

So, everybody will need (A) continuously consider both sides of the same coin, as well as related human-centric, societal, sustainable, economic, and other perspectives, and (B) to identify, deploy and continuously monitor and optimise the appropriate contextual, nuanced symbiosis of these essential components, dimensions, and perspectives.

This blog aims to provide further guidance in making AI truly work, not only function.

The blog is part of a Series, so if you have not already done so, please also read Part 1 of this Series where we introduce the notion that human-centred AI can become an enabler and facilitator for the climate of change we need in Europe, and worldwide.

Digital Ecosystems Are Interconnected Vessels

Digital has become a must-have, for people, society and economy. Digital platforms, AI, robotics, edge computing and the internet of things (IoT) are further expediting this process by connecting, inter-connecting respectively hyper-connecting individuals, organisations, communities, societies and data, with tens of billions of objects and entities.

All these technical capabilities and related digital ecosystems generally comprise of a technical stack that to some extent can be visualised as set forth below in Figure 1. These are made up of some combination of the various forms of data together with software-enabled algorithms that have sufficient computing power either centralised, decentralised or distributed on the Edge or in IoT devices, and interfaces, connectivity and infrastructure where necessary.

Digital Ecosystems: Technical Stack + Data

Figure 1: Digital Ecosystems Are Interconnected Vessels

Making Technology Function is Difficult Enough

When preparing the relevant kitchen tools, cooking ingredients, basic cooking skills and a plan what to cook, one can come up with the technical functions, and the functional specification, technical requirements, the technical specification, and thereafter the actual development and engineering. Right after, it is time to demonstrate it functions, and one is all set. Right?

We all know it is difficult enough to make technology function. Especially regarding AI, making it function is not an easy feat.

AI technology is an inherent component of Industry 5.0. However, even if the technology itself may be at the right technical readiness level, the readiness of a technology on itself does not guarantee its success.

Although new and seemingly burdensome for some, it will for sure be beneficial in order to truly make it work, with AI in the equation. Before one notices, it will become second nature. The ‘it’, in ‘make it work’ is not AI or other technological functionalities or capabilities; it is a valued use case that addresses Societal Challenges of any kind.

Risk in Cyber-Physical and other Digital Ecosystems

When thinking and talking about risk, it is important not to see risk as something necessarily negative.

It is an integral part of the equation and with that an enabler and facilitator of anything that works in a trusted, trustworthy and accountable way. It gives essential and valuable insights in what may happen or may go wrong, what people or society like or fear, etcetera. For sure, in the AI or AI-supported domain that is an essential success factor.
The magnitude of risks, determined by the probability as well as the impact thereof, is very much context and application dependent. To prepare for and mitigate the potential harm, to embed preparedness for foreseen and unforeseen situations, and to make it resilient and future-proof, it is necessary that AI systems are designed and deployed guarded by trust principles. These non-functionals are principles that consistently preserve trust, trustworthiness and engagement of all relevant stakeholders. Examples of such principles are security, safety, privacy, transparency, auditability, sustainability and robustness. There are several hundred of trust principles. These can be found in best practices, guidelines, white papers, standards, regulations but also in common practice and nature.

Two major challenges in the AI design and deployment are (1) to map the relevant risks accurately and comprehensively throughout the system’s entire lifecycle, and (2) to incorporate non-functionals by design.

**Risk Segmentation; Creating Insights & Oversight**

Risk is not a four-letter word, and – even in the AI context – deserves its own series of studies, publications and the like.

In any case it is useful to segment the various AI-related dimensions of this Digital Age in order to get some relevant oversight and insight. Segmentation provides structure, insight and oversight, and facilitates awareness, understanding and appreciation.

Keeping the holistic, end-to-end ecosystem mindset and approach, an initial segmentation into four (4) categories can be done: Non-connected, Connected, Inter-connected and Hyper-connected.

- **A. Non-connected**, which is a stand-alone device, tool, machine, appliance or application that does not have connectors or connectivity that can connect to the internet or other external network or resources.
- **B. Connected**, where a device, tool, machine, appliance, application or system may be connected to, via the internet, a centralised database, cloud infrastructure and other centralised systems;
- **C. Inter-connected**, where several edge devices, tools, machines, appliances, applications or systems are connected with each other, either via orchestrated, federated systems, and;
- **D. Hyper-connected**, where numerous far edge and other IoT devices, tools, machines, appliances, applications or systems are directly connected with each other via distributed (computing and related) ecosystems of ecosystems.

For each of these segments, various value cases, business models, feasibility models and therefore use cases can be identified and created in the AI-supported Industry 5.0 domain. Each segment has its own values, benefits, efficiencies, inefficiencies, etc.

The segmentation set above obviously is not the only one possible. Various other segmentations are to be relevant to considered as well, such as for instance real-time, near-real-time or not. This segmentation may be relevant when near-real-time autonomous 3D printing is considered, or real-time prognostic health monitoring or related integrated logistics support is relevant. Other segmentations that can be considered are single-vendor, multi-vendor, OEM, public, private, public-private, etc.

**Risk Classification Spectra: A Multi-Layered Approach**

When focussing on one of the above-mentioned segments, Hyper-Connected devices, and taking a risk-perspective to those, a methodology to do high-level quality risk classification is to have a multi-layered approach and do such risk classification per spectrum, starting with the risk classification of the connectors and connectivity of the IoT device itself.

Even though AI capabilities may not yet be in the equation, it is essential to understand the various risks that are embedded in or could arise from such IoT device. Subsequently, other risk spectra should be considered and risk classified, as visualised below in Figure 2.

Especially more downstream there may be risk spectra that may not be relevant; however, if such spectrum may become relevant later in the life cycle of the IoT device, it is recommendable to keep it in and already do the spectrum risk classification. In general, three categories of main risk levels are used: low, medium and high. Based on the outcome of (i) a risk classification for each spectrum, and (ii) the interim outcome of the various risk classifications up to Risk Spectrum layer 13 (AI Capabilities), the baseline risk classification can be established.

Based on that baseline, the AI Capabilities risk classification can be done, and the subsequent risk spectra; the holistic perspective constitutes the Combined Risk Classification, on which one can consider and organise technical & organisational security, safety, privacy and related technical and organisational measures.

In any case, the segments, whether non-connected, connected, inter-connected or hyper-connected, that have AI capabilities of any kind, are for sure game changing, where non-functional and functional requirements have to be addressed together.

The winner will be the one who understands fully the societal challenges at hand and related sectoral requirements.

**Success By Design**

Non-functionals are as important as functionalities. Even better, they positively augment each other if balanced out intelligently and correctly. The symbiosis of both is a main success factor for any development and deployment of AI.

For sure, Industry 5.0 and related ecosystems, including the persons, organisations and other stakeholders therein, can benefit from this, and can improve itself towards human-centric, secure, safe, sustainable, trusted, trustworthy, resilient and otherwise future-proof systems.

How to methodologically make that work? We will discuss various proven methodologies in our subsequent blogs, so please stay tuned.
Manufacturing Data Spaces for Artificial Intelligence: the EC perspective

By: Andrea Ferretto Parodi, Matteo Falsetta / GFT Italy

It is a commonsense statement that the quality of advanced AI-based decisions heavily depends on the quality of the data we introduce into the system: Garbage in, Garbage out. Common European Data Spaces are currently discussed, developed and deployed in several initiatives at European level with the aim to break the silos and fill the gaps that currently prevent a pan-EU “free flow of data”. In the Manufacturing sector, four initiatives are to be considered in order to determine a state of play regarding this strategic matter: The EC directives and communications; the BDVA (Big Data Value Association) working groups; the IDSA (International Data Spaces Association) publications and papers; the Connected Factories (EFFRA Euro- pean Factories of the Future Research Association) and Digital Manufacturing Platforms ecosystem. In this first blog regarding Manufacturing Data Spaces for AI, we will address the EC viewpoint through its communications and recommendations.

The 2018 EU Data Package23 included a specific staff working document on how to implement data sharing spaces in the Private Sector24. Three main Business Models have been identified for the Private sector to fully benefit from the Data Economy and the Data Revolution: the Open Data model, the Data Monetisation model and the Trusted Ecosystem model.

In the first model, private sector companies could contribute to the Open Data movement by disclosing and publishing Datasets which do not include privacy or confidentiality issues. The typical case includes out-of-production data maybe aggregated, anonymised or pseudonymised, put at disposal of open innovation “living lab” ecosystems for advanced experiments. More recently, this has become an important topic for the so-called Teaching / Learning / Didactic Factories where companies and students can find a hands-on “test before invest” facility to innovate and experiment.

In the second model, private sector companies are able to valorise the data they produce, to provide an internal or externalised Data Marketplace and to start and develop new servitisation business models. Manufacturing enterprises, e.g. in the sector of Machine Tools and Robots, publish and valorise high value Datasets which could be used by service providers to develop and test their advanced service value propositions, e.g. in diagnosis and maintenance of complex equipment. The monetisation of such a business could be implemented in different ways like credits or virtual coins and not just by immediate e-commerce and payment functions.

In the third model, private sector value chain is organised in hierarchical (e.g. tier 1, tier 2, tier n) or non hierarchical (e.g. SMEs ecosystems) trusted networks implementing in a more flexible and configurable way what the EDI did many years ago. Hierarchical trusted network usually follows the business and governance model of a large company (e.g. a car manufacturer, a ship building, a food producer) which dictates the way the network should behave in order to contribute to the trusted cluster. Along the time, such rigid models and chains have been gradually transformed in more open networks where entry barriers have been substantially lowered and internal competition rules blurred. Non-hierarchical trusted networks are usually implemented by local regional districts of SMEs usually sharing in real time their productive capacity and allow dynamic, on-the-fly matching between demand and offer of manufacturing services (e.g. MaaS Manufacturing as a Service).

The 2020 EU Data Strategy25 finally defined four main pillars in order to implement Data Economy: A cross-sectoral governance framework for data access and reuse (now implemented by the Data Governance Act), a set of Technology Enablers implementing Personal and Industrial Data Platforms, a pool of new professions and competencies to introduce Data Economy in the enterprise, an ecosystem of common EU Data Spaces in several crucial economic sectors, such as Manufacturing.

The STAR ICT-38-2020 project is one of the H2020 projects aiming at applying the EU Data Strategy principles to the field of “Data for AI in the Manufacturing sector”, especially focusing on technological enablers and pillars for Data Quality, Cyber Security, Explainable and Trustworthy AI. Moreover, STAR technology assets enable the three Business Models of the 2018 Data Package through its Virtual DIHs (Open Data), Assets Catalogue (Data Marketplace) and its three industrial pilots (Trusted Networks) in the domains of Human-Cobot Collaboration for Robust Quality Inspections, Human Centred AI for Agile Manufacturing 4.0 and Human Behaviour Prediction and Safe Zone Detection for Routing.

---
