OPEN ACCESS
IOP Publishing | ISECAEP
Journal of Reliability Science and Engineering
J. Reliab. Sci. Eng. **1** (2025) 015001 (9pp)
https://doi.org/10.1088/3050-2454/ad9f63

Opportunities and risks of artificial intelligence for industry 5.0 in the context of reliability and maintenance engineering

Michele Compare[1] **and Enrico Zio**[1,2,3,*]

[1] Aramix s.r.l., Milan, Italy
[2] Department of Energy, Politecnico di Milano, Milan, Italy
[3] Mines-Paris, PSL University, CRC, Sophia Antipolis, France

E-mail: enrico.zio@polimi.it

Received 14 October 2024, revised 9 December 2024
Accepted for publication 9 December 2024
Published 16 January 2025

Abstract

The Industry 5.0 (I5.0) paradigm is expected to further boost the relevance and widespread application of Artificial Intelligence (AI). The actual contribution that it can bring is challenged by current technology, scientific contexts and trends. The objective of this work is to provide an overview of some of the issues that need to be tackled for AI to support the development of reliability and maintenance engineering for I5.0. Three use cases of AI development opportunities are discussed, which are fully compliant with the EU vision of AI enhancements for I5.0: lumping together data of different origins and scales, expert knowledge combined with AI, and causality-based AI. These use cases show that there is room for advanced and successful AI solutions for I5.0, and allow identifying three main elements required to grasp opportunities while identifying, preventing and mitigating risks related to AI development: multidisciplinarity, experience and continuous training.

Keywords: artificial intelligence, industry 5.0, maintenance

1. Introduction

Industry 5.0 (I5.0) is characterized by strong human-machine cooperation and commitment to environmental sustainability. Its main goal is to bring added value to production through highly customized products that meet the specific needs of consumers (e.g. [1, 2]). According to the EU strategy, there are four main pillars for the transition towards I5.0 ([2–5]):

- Human-centric technologies: up-skilling and re-skilling of workers, especially in digital skills, and adoption and development of technologies that adapt to the workers, rather than the other way around.
- Competitiveness: research and innovation in advanced technologies (e.g. robotics, IoTs, 3D printing, cloud computing) to increase productivity, optimize processes, and remain competitive in the global market.
- Sustainability: reducing the environmental impact of activities with renewable energies and eco-friendly production practices towards a circular economy.

* Author to whom any correspondence should be addressed.

- Resilience: increase robustness and flexibility in industrial production and business processes, to cope with disruptions and crisis (e.g. geopolitical shifts and natural crises, such as the Covid-19 pandemic).

I5.0 represents the natural evolution of Industry 4.0 (I4.0): whilst the objectives of the latter were digitalization and further automation, enabled by the development of increasingly advanced technologies, the focus of the former is on leveraging the results of I4.0, (e.g. ubiquitous presence of increasingly powerful Internet of Things devices and more advanced Cyber-Physical Systems) for the development of solutions that make production more sustainable, resilient and competitive on a long-term basis. The solutions developed must be grounded within a more human-centric perspective, which considers at least three different levels: i) consumers, who expect personalized products on a broad scale; ii) appropriately skilled workers, who interact with robots; and iii) actors in global challenges for the survival of the human species (e.g. climate change, ecosystem collapses, pandemics, etc) who can adopt and promote more sustainable lifestyles through all daily decisions.

AI plays a fundamental role in achieving the goals of the I5.0 paradigm shift. In particular, Human-Centered AI is asked to bridge the gap between AI and Intelligence Augmentation (IA), i.e. between automating processes and augmenting the capabilities of the human in the process [6]. This offers great opportunities for scientific and technological enhancements, but also poses new challenges to the competitiveness of the solution developers, with consequent new risks that must be evaluated and managed (e.g. [7]).

The aim of the present work is to indicate some elements of competitiveness in AI solutions for I5.0 and some risks; then, this contribution is not intended to be a literature review of the AI algorithms applied to reliability and maintenance engineering (the interested readers can refer to very recent works covering this topic, e.g. [8–10]). We start from sketching the current technological and scientific contexts in section 2. To present some opportunities related to the development of AI solutions within the sketched contexts, in section 3 we propose use cases pertaining to reliability and maintenance engineering, two central topics to meet the sustainability goal of I5.0. The associated risks are discussed in section 4. Conclusions are drawn in section 5.

2. Context

In the last five years, there has been a proliferation of algorithm libraries, developed particularly under the direct control of the 'over the top' companies (Google, Meta, Microsoft, etc). For example, TensorFlow and Pytorch, which are among the most popular software frameworks for AI development, have been developed and are maintained by Google and Meta, respectively ([11, 12]). In many cases (e.g. generative and multimodal AI), algorithms are pre-trained using huge computational and economic resources and made available on the market already as ready-to-use. The huge size of these models (i.e. the number of parameters they are based on) makes their retraining or customization very complex, from both technical and economic perspectives [13]. On the one hand, this situation has resulted in a widespread ability to develop solutions based on AI algorithms and models: roughly, all users have access to the same models and technologies, generally under the same economic conditions. On the other hand, it has made it difficult to clearly explicit the differentiating elements of AI solutions, especially for I5.0. In fact, the AI market opportunities are not so much in the algorithms themselves but in the ability to use the appropriate ones and run them on quality data, in a cyber-secure environment [14].

A relevant example is given by the surge of solutions based on Foundation Models (FMs), a general class of models trained on broad data (generally using self-supervision at scale) that can be adapted (e.g. fine-tuned) to a wide range of downstream tasks [15]. FMs, whose name highlights their critically central yet incomplete character, include Large Language Models (LLMs), which are used also for I5.0 applications (e.g. expert systems for maintenance based on conversational chatbots trained on manuals, maintenance reports, etc). Of these, there are several models available for free use (Alpaca, Vicuna, ChatGPT3, etc [16–19]), pre-trained using significant resources. For example, ChatGPT3, with 175 billion parameters, was trained on Microsoft Azure AI supercomputer with estimated cost of US $12 million [20]. The most performing FMs, i.e. with huge numbers of parameters, are however linked to commercial offerings (e.g. ChatGPT4, Gemini, Claude [21, 22]). Independently on the size, these models all use transformer architectures ([23, 24]) and the attention mechanism ([25]), which have greatly improved the ability of language models to handle long-range dependencies in natural-language text.

Various companies offer solutions based on pre-trained models, especially through the Retrieval-Augmented Generation (RAG, [26, 27]) approach, which conveys the LLM capabilities to specific domains and internal knowledge of an organization without the need to retrain the LLM, while ensuring that the output remains relevant and accurate. For these solutions, however, it is difficult to technically outstand others, thus leaving competitiveness to commercial aspects, as witnessed by the increasingly economical offers of these services.

There has also been an exponential growth in the technical and scientific literature on AI algorithms, with articles easily accessible online and often for free (open access), given the widespread success of open repository publication models (e.g. arxiv.org). Coherently with this, in the present work we have selected references that are all open access.

This extreme knowledge-sharing on the one hand brings benefits to the technical development and scientific advancement of AI; on the other hand, it increases the difficulty in selecting scientifically relevant advancements, especially for research outside the mainstream, as in the case of niche

042

industrial applications for which there are few reference experiences to compare with.

The process of disseminating and sharing theoretical knowledge on AI algorithms and models, known as AI democratization, can lead to stagnation, if not decline, in the diversity of AI research: some of the largest and most prestigious universities competing for the leadership in AI, mainly based in the US (MIT; University of California, Berkeley; Carnegie Mellon; Stanford University) and in China (Chinese Academy of Sciences, [28]), have much lower thematic diversity in AI research than expected from their volumes of activity and public nature [29]. These influential universities end up to be strong collaborators of large private companies, leading to a certain homogenization at the top of AI research. In this respect, it has been estimated that almost 80% of AI research is on resource-intensive development of FMs [30]. This results in some thematic areas and application domains, particularly those of I5.0, being not sufficiently covered [31].

In this scenario, some distinctive elements that AI players can adopt to offer competitive solutions to I5.0 are:

- Systemic vision: to add value to an AI-based solution, it is essential to consider and know the industrial context in which it is intended to work. A multidisciplinary approach to the development and implementation of the solution is needed, where AI skills are supported by software development, specialized engineering and business skills for meeting the industrial requirements of the I5.0 paradigm.
- Experience. The availability of different solution alternatives would require, to consider them all, too long time spans for development, adaptation and performance analysis of each alternative compared to the time and budget of the industrial context. Although AutoML approaches (e.g. [32–34]) and large FMs help speed up this process, it is undeniably valuable to be able to evaluate the appropriateness of available algorithms and models in relation to the knowledge, information and data available, the scientific rigor and the specific needs of the use case. For example, data may not always be available for certain solutions, and algorithms do not always produce the results reported in the literature when applied to different data. In practice, the most innovative algorithm is not always the most suitable to the context.
- Identification of AI research, development and innovation areas capable of responding promptly and with the best technologies to specific I5.0 needs not yet fully intercepted by commercial solutions. This requires highly advanced technical and scientific skills, and constant attention and investment in research and innovation.

These competitiveness elements fully match with the EU vision of AI for I5.0, which pushes the AI enhancements in several regards [4], including causality-based and not only correlation-based AI (e.g. [35]), handling correlations among complex, interrelated data of different origins and scales in dynamic systems of systems (e.g. [36]), informed deep learning (i.e. expert knowledge combined with AI, e.g. [37, 38]).

3. AI Development Areas

In this section, we discuss some research and development opportunities for the AI enhancements discussed in section 2, consistently with the elements of competitiveness discussed therein. These opportunities relate to reliability and maintenance engineering, which play a central role still in I5.0, as they underpin main I5.0 pillars such as quality, waste reduction, safety of workers and consumers, etc.

Specifically, we consider three different use cases: handling diverse data from systems of systems, expert knowledge combined with AI and causal modeling.

Other AI research pathways in support to I5.0 identified by EU (i.e. Swarm intelligence for robotics, Brain-machine interfaces, person-centric AI, Secure and energy-efficient AI), pertain to more technology-specific issues, which are not considered in the following analysis.

3.1. Handling diverse data from systems of systems

In spite of the relevance of FMs in all contexts of our society, their application to I5.0 currently is limited to providing conversational support to query large company documentation or automate some documentation tasks. To the Authors' best knoweldge, this is corroborated by the evidence that when querying Google Scholar with any combinations of words 'Industry 5.0' 'LLM/Large Language Models', 'GPT' only a few papers are found, proposing LLM for chatbots and for interacting with industrial robots or autonomous vehicles (e.g. [39]). Paradoxically, the limitation of use of LLM in I5.0 applications is supposed to be the exact objective of sustainability, given the open question about the energy sustainability of this technology [40].

A different perspective to effectively exploit FMs for I5.0 applications is to embed them in projects, which need to leverage past field experience, available through data of different origins and scales, stored in different databases.

Just as an example, consider a use case related to the design of the production process in a mechanical components manufacturing company: past solutions form the basis for estimating production KPIs and, thus, production costs. However, the accumulated experience is typically available in different forms: drawings, numerical data, technical reports, etc. Yet, the amount of data is surely not enough to train an FM. To simultaneously process and interpret multiple types of data, multimodal AI solutions are needed, offering a more comprehensive understanding of its inputs, leading to more accurate and context-aware responses. To this end, we have developed a solution that integrates (figure 1):

- A tool for extracting relevant features from industrial drawings. This is based on Gemini, with prompting optimized to focus on information relevant to the designer's interest, which is related to the difficulty of producing a component with given geometrical characteristics.
- The database related to the operational profile of the machines used to produce components in the past, to characterize working conditions.

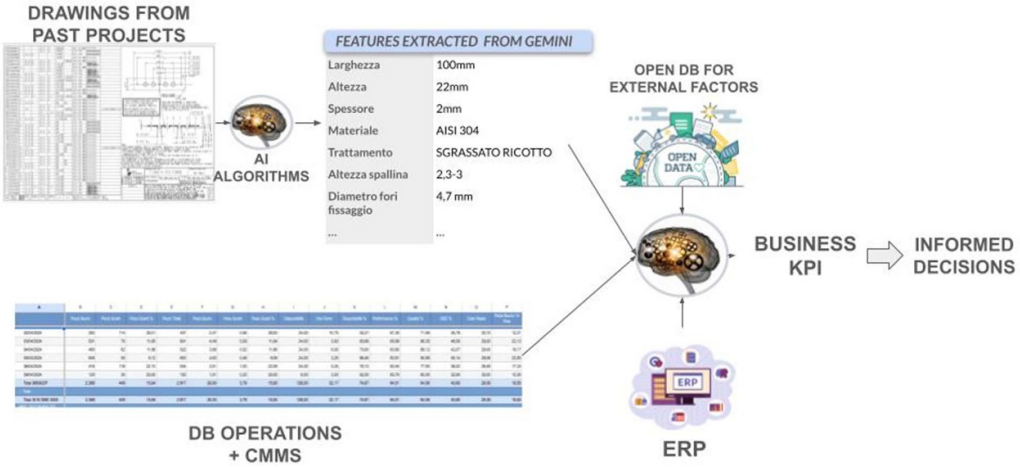


Figure 1. Multimodal Approach.

- The company's Enterprise Resource Planning, to consider timings, costs, and other production KPIs of the components produced in the past on a specific machine.
- Open Data to include exogenous variables that may impact performance (e.g. external temperatures affecting the melting phase).

It is from the combination of all these pieces of information that we can predict the KPIs of production of a new component on a given production machine, and on this basis find the optimal allocation of components to production machines.

3.2. Expert knowledge combined with AI

The availability of limited datasets is quite widespread in practice, even in I5.0 settings. For example, for training algorithms for predictive maintenance applications with good generalization properties, we need datasets on the degradation and failure behaviors of similar systems. In many cases, to gather datasets with appropriate informational content we need to consider a fleet of (similar) systems over a sufficiently long operational life, before the obsolescence of the systems themselves, or their revision, or significant revamping. With respect to this possibility, it is clear that there can be a significant difference between Small and Medium Enterprises (SMEs) and large companies, in relation to the scalability of the solutions.

On the other hand, many AI algorithms, which much of the academic and industrial AI developments focus on, are actually Machine Learning or Deep Learning algorithms. These algorithms struggle with applications for which the amount or quality of the available data is limited.

Then, a distinctive element in proposing effective solutions for I5.0 is decision support under this condition of limited data, using AI algorithms.

As an example, consider a use case that requires decision support to define the tests that are usually performed by manufacturing companies to estimate the reliability of the components produced. These tests are generally long and expensive and, then, are carried out by accelerating the degradation processes of the components [41]. Accelerated life test models are used to define the relationship between accelerated and normal conditions and to translate the accelerated results to the expected real life performance. The objective of acceleration is the minimization of the number of test hours to be achieved by setting the test conditions (i.e. forcing degradation stresses). For this, it is also essential to exploit a-priori engineering knowledge on degradation mechanisms and accelerating factors.

For example, we can rely on the framework developed by Foster ([42–45]) for optimal reliability parameter inference. A Bayesian perspective is used to leverage the expert knowledge on the component reliability characteristics for setting relatively narrow a-priori distributions on the model parameters. Specifically, let θ denote the latent variables to be learned from the test (e.g. Weibull distribution parameters and Arrhenius acceleration model parameters), and let ξ represent the experimental test setting (i.e. the stress variables defining the acceleration). By introducing a prior $p(\theta)$ and a predictive distribution $p(y|\theta,\xi)$ for the test outcome y, we can estimate the Expected Information Gain (EIG, [46]):

$$\mathcal{I}(\xi) = \mathbb{E}_{p(y|\xi)}\left[H[p(\theta)] - H[p(\theta|y,\xi)]\right] \tag{1}$$

where $H[\cdot]$ represents the entropy of the argument distribution and $p(\theta|y,\xi) \propto p(\theta)p(y|\theta,\xi)$. In words, we seek for the experiment that is expected to bring the largest amount of information. This allows deriving a posterior distribution with a smaller entropy value, corresponding to a larger gap with respect to the prior distribution entropy. The test design process now amounts to finding the test setting ξ^* that maximizes $I(\xi)$.

By repeatedly applying Bayes Theorem, we can prove that [42]:

$$\mathcal{I}(\xi) = \mathbb{E}_{p(y|\xi)}\left[\mathbb{E}_{p(\theta|\xi,y)}\left[\log p(\theta|\xi,y)\right] - \mathbb{E}_{p(\theta)}\left[\log p(\theta)\right]\right] \tag{2}$$

$$= \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\log \frac{p(y|\theta,\xi)}{p(y|\xi)}\right] \tag{3}$$

$$= \mathbb{E}_{p(\theta)}\left[H[p(y|\xi)] - H[p(y|\theta,\xi)]\right] \tag{4}$$

$$= I_{KL}(\xi) = \mathbb{E}_{p(y|\xi)}\left[\mathrm{KL}\left(p(\theta|\xi,y)\,\|\,p(\theta)\right)\right] \tag{5}$$

044

where $KL(p\|q)$ is the Kullback–Leibler distance of distribution p from distribution q [47]. In particular, equation (4) can be intuitively interpreted as follows [48]. When maximized, the first term prefers examples ξ for which there is uncertainty in the predicted outcome y. Using this as a selection criterion is equivalent to uncertainty sampling: pick the value where there is more uncertainty to reduce it through the experiment. However, uncertainty sampling can have problems with examples that are inherently ambiguous. By adding the second term, we penalize such behavior, as we add a negative weight to points whose predictive distribution is entropic even if we know the parameters. Estimating $I(\xi)$ for a single test is computationally burdensome, even in the simplest settings [42], like the Prior Contrastive Estimation (PCE) lower bound on $I(\xi)$:

$$I(\xi) \geqslant \mathbb{E}_{p(\theta_0)p(y|\theta_0,\xi)p(\theta_1)\ldots p(\theta_L)}\left[\log \frac{p(y|\theta_0,\xi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(y|\theta_\ell,\xi)}\right]. \tag{6}$$

Efficient estimation can be done using finite samples by:

$$\hat{I}_{\mathrm{PCE}}(\xi) = \frac{1}{M}\sum_{m=1}^{M} \log \frac{p(y_m|\theta_m,\xi)}{\frac{1}{M}\sum_{\ell=1}^{M} p(y_m|\theta_\ell,\xi)} \tag{7}$$

where $y_m, \theta_m \sim p(y,\theta|\xi)$.

This bound is used to optimise ξ by a stochastic gradient process [42].

With respect to the reliability testing, assume that the distribution of the Time to Failure is the Arrhenius Accelerated Life Model embedded in the Weibull distribution. The hazard rate over time t reads (e.g. [41])

$$h(t|\xi;\theta) = \frac{1}{\alpha}\cdot\frac{1}{e^{\beta^T\cdot\xi}}\cdot\left(\frac{t}{e^{\beta^T\cdot\xi}}\right)^{\frac{1}{\alpha}-1} \tag{8}$$

where ξ represents the settings of the Φ accelerating factors, $\theta = (\alpha;\beta)$, where $\alpha \in \mathbb{R}_+$ is the shape factor of the Weibull distribution and $\beta \in \mathbb{R}^{\Phi}$ is the weighing vector.

In a very simple case, we assume that we have ten components to test up to failure (i.e. no censoring), $\Phi = 2$, $\alpha \sim Ga(\alpha|2,2)$, $\beta_k \sim Ga(\beta_k|2,2)$, $k = 1,2$, $\xi = [0,10]^{\Phi=2}$. As it emerges from figure 2, the result of the optimization of test setting ξ is to run five components at the maximum level of one accelerating parameter with no acceleration of the other parameter (e.g. points around (10,0)), whereas the remaining five components are tested in the dual configuration of parameters acceleration (e.g. points around (0,10)).

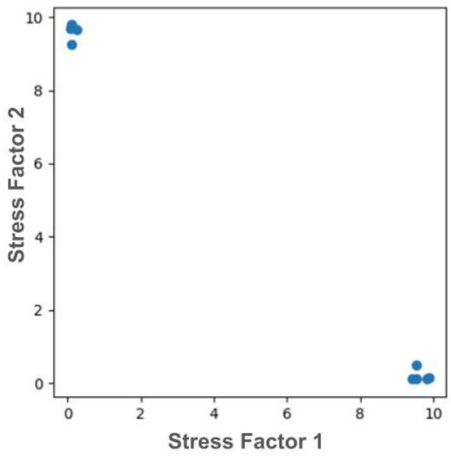


Figure 2. Application of Bayesian Optimization for the design to reliability testing.

3.3. Causal modeling

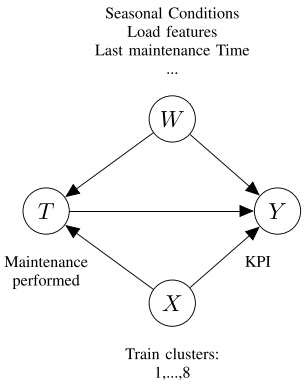
According to the EU vision, another area in which AI can bring added value to I5.0 when combined with engineering experience and knowledge is that of causal analysis. This focuses on the cause-effect relationship among variables or events in physical, behavioral, social and biological sciences, evaluating explanations for an observed scenario and predicting the effects of actions and policies on scenario development. For a deep dive into the basics of Structural Causal Models (SCMs), the reader is referred to [35, 49, 50].

Formally, an SCM is defined as a triplet $M = \langle U, V, \mathcal{F}\rangle$ where:

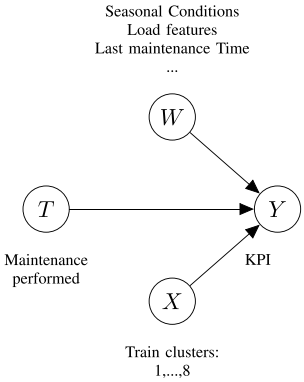
- U is the set of background variables, representing unmodeled causal influences.
- $V = \{V_1, V_2, \ldots, V_N\}$ is the set of variables of interest, influenced by $U \cup V$.
- $\mathcal{F} = \{f_1, f_2, \ldots, f_N\}$ is the set of functions such that $v_i = f_i(pa_i, u_i), i = 1,\ldots,N$, where pa_i represents the set of parent variables of V_i and u_i represents the set of background variables on V_i. The functional form of f_i can characterize any stochastic mapping.

Any SCM can be associated with a corresponding Directed Acyclic Graph (DAG) $G(M)$, whose nodes correspond to variables U and V, whereas directed edges link U_i and pa_i to V_i, $i = 1\ldots,N$. Every parent is a direct cause for all its children.

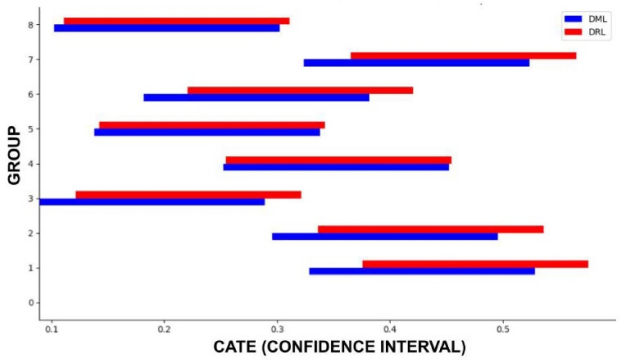
The DAG in figure 3(a) represents an SCM relevant to a maintenance application. As usual, for clarity the variables U, one for each node V, are not explicitly shown in figure 3(a). The model is used to evaluate the effectiveness of the different interventions performed on a specific subsystem of a fleet of trains, which have been clusterized into eigth groups based on features related to their operation and registry data. The effects of maintenance on KPIs linked to system availability are very different, both when comparing different types of intervention (e.g. lubrication, balancing, replacement, etc) and considering the same type of intervention at different times. This difference in performance is possibly due to different confounders such as seasonal conditions, past maintenance interventions, loads. These factors can modify the causal link between treatment (maintenance action) and outcome (KPI value). For example, the seasonal conditions confounder is introduced to model the fact that seasonality aspects (i.e. harsh weather conditions) might affect both the capability of the operator to perform the procedure and the following loads on systems, which the KPIs depend on.



(a) Causal Model



(b) DO calculus



(c) CATE of Lubrication estimated with DML (blue) and DRL (red)

Figure 3. Causal Modeling.

The SCMs can be used for probabilistic inference. In this case, they can generalize Bayesian Networks. However, each SCM naturally defines a qualitative hierarchy of concepts, described as the 'ladder of causation' [51]. The associational level for probabilistic inference is at the basis of the ladder, where we use statistical learning to try to infer properties of dependence among random variables from observational data. The other two levels are interventional and counterfactual. Each level corresponds to a distinct notion within human cognition and allows formally articulating qualitatively different types of questions regarding the observed variables of the underlying system [52]. Causal reasoning concerns these levels [50].

Specifically, the interventional level concerns actions taken to modify the values of some variables and are used to establish a causal relationship between the manipulated variables and the outcome of interest. This allows understanding the results of specific actions or changes.

Formally, interventions are denoted using the *do*-operator and define a submodel of SCM $M = \langle U, V, \mathcal{F} \rangle$. For a set of variables $H \in V$, we can consider a realization η and the submodel $M_\eta = \langle U, V, \mathcal{F}_\eta \rangle$, where $\mathcal{F}_\eta = \{f_i : V_i \notin H\} \cup \{H = \eta\}$.

With respect to the reference example, we might be interested in the effect of a specific treatment T. This formally reads

$$P(Y|\mathrm{do}(T)) \neq P(Y|T). \tag{9}$$

Practically, to build M_η, $H = T$ and $\eta = \mathit{Lubrication}$, we delete the links $W \longrightarrow T$ and $X \longrightarrow T$, set T and replace equation $T = f_T(W, X, u_T)$ with $T = \eta$ (figure 3(b)).

Counterfactual reasoning, the third level of the ladder, assumes that each variable can be given hypothetical or unobserved consequences that would have occurred under different treatment or intervention conditions. The basic idea, thus, is to ask what would have happened in a situation had certain things been different. This is like rewinding the world, changing a few crucial details and, then, predicting what happens in the fictional world. By tweaking the right variables, it is possible to separate true causation from correlation and coincidence.

Formally, given a Probabilistic causal Model (PCM) $\langle M, P(u) \rangle$ and some evidence $E = e$ (i.e. observable variables set to values e), the probability $P(v_{jv_i}|E = e)$ of the counterfactual 'given $E = e$, if it were $V_i = v_i$, then $V_j = v_j$' is computed through three steps:

1) Abduction: Bayesian update $P(u|e)$, as usual in Bayesian frameworks
2) Action: build submodel M_{v_i}
3) Prediction: compute $P(v_{jv_i}|e)$ through PCM $\langle M_{v_i}, P(u|e) \rangle$.

In the maintenance example, counterfactual reasoning allows answering questions such as: 'what would have happened if instead of carrying out the rebalancing maintenance intervention, one had carried out the lubrication intervention given that the balancing actually caused the unavailability of the system after eight months in the operating context defined by the operational variables W?'.

Given the SCM, the quantitative identification of the link between each type of intervention and the effect it produces, expressed by the average conditional effect of the treatment (conditional average effect of the treatment, CATE) offers the possibility not only of identifying the best intervention in the operational context that occurs, but also to retrospectively analyze the decisions made for improvement actions. This can be done through different techniques, for example the Double Machine Learning and Doubly Robust Learning [48] (figure 3). From this, we can argue that in the case considered there is always a positive effect of lubrication, although with different results depending on the type of

046

train. This information is useful to perform further analysis to explain the difference.

4. Risks

Although there are many opportunities of applications of AI to I5.0, there are also relevant risks, which can be classified in data risks, model risks, operational risks and legal risks ([53]). Additional references on specific risks can be also found in [15, 54–56]. However, the objective of this Session is to share the risks that are deemed by the Authors as more relevant for the AI diffusion in the reliability and maintenance engineering context, based on their experience in the field.

Traditional software development uses structured methodologies that allow detailed planning, and Verification and Validation (V&V). On the contrary, the development of AI models is characterized by an iterative and experimental approach, which the 'trial and error' part is inherent to. Although this does not mean significant change in the phases of software development (design, development and release into production), nonetheless it poses risks with regards to the execution times of project development and, above all, makes it difficult to offer the industrial client guarantees on the results of the solution, as these depend on the data available, in terms of quantity and quality, on the problem considered, etc. For this reason, the development of AI solutions for the industrial sector often starts timidly with Proof of Concept activities, for which guarantees on the results are not offered, generally at economic conditions that make this phase not very expensive from the investment perspective from all involved stakeholders. The commercial risks with respect to this model are evident: effort is spent on activities with generally low added value and which, therefore, risk to be left without continuation.

Another relevant risk can be found in the gap between the advanced mathematical, IT and engineering skills necessary to develop effective AI solutions, and those of the industrial context, especially in SMEs. This gap can lead to misunderstandings of project demand and response, with the risk that even in cases of excellent outcomes, the final solution is not given its proper value. One factor that contributes to this is that traditional software is based on algorithms and rules that produce outputs that are repeatable and interpretable, whereas AI models generate probabilistic and in some cases counter-intuitive responses. It is, therefore, important to also work on training end users for increasing their awareness and confidence in the use of AI solutions.

A further risk relates to the speed of evolution of AI algorithms and their rapid obsolescence, even when hyper-customized to the user needs. In this regard, consider how the LLMs have quickly made obsolete some features of many process management softwares (for example those based on bots, Optical Character Recognition, etc). As a consequence, the fast-evolving technology on the one hand results in the need to invest in solutions that have a very short return on investment, and on the other hand, this requires the capacity for continuous adaptation to algorithmic evolutions.

Finally, regulatory compliance can be a particularly subtle risk given the diversity of regulatory frameworks, the speed of change that characterizes the current phase and the abundance of vague principles (e.g. the Italian AI Strategy document [57] gives statements like 'AI should be developed, adopted, and used in a human-centered, trustworthy, and sustainable way'; 'It is essential that these systems are used safely and that citizens and businesses can trust that they respect fundamental rights, such as privacy and data protection, and can be used without risk of bias and discrimination'). The regulatory evolution expected in the coming years certainly represents one of the issues to be managed. This uncertainty significantly hampers investments in AI for industry.

5. Conclusions

The diffusion of AI and the rapidity of its evolution bring many opportunities, challenges and risks, not only technical, but also economic, social, ethical and anthropological. This is typical of technologies that have not yet reached a high level of maturity, and makes entrepreneurial, scientific and technical activities very complex in the balance of opportunities and risks of developing AI solutions, particularly in I5.0.

The objective of this contribution was to identify and discuss some opportunities and risks for awareness and informed decisions. These opportunities have been shown with respect to the reliability and maintenance issues, which remain central in the I5.0 paradigm. Specifically, we have proposed three use cases of AI development opportunities, which are fully compliant with the EU vision of AI enhancements for I5.0: lumping together data of different origins and scales, expert knowledge combined with AI, and causality-based AI. These use cases have shown that there is room for advanced and successful AI solutions for I5.0, provided that these build on the three elements identified to grasp opportunities while identifying, preventing and mitigating risks: multidisciplinarity, experience and continuous training.

Specifically, when we need to use different types of information sources, as in the case of manufacturing of mechanical components, it clearly emerges that a systemic vision and a sound experience are a plus for the successful project delivery, as this requires expertise in industry and business, as well as in AI and data science, together with a deep knowledge about the latest technologies, e.g. for effectively prompting multimodal FMs. This need of lumping together different pieces of information asks for further research work aimed at developing techniques and methodologies capable of dealing with relatively small sets of data, as this is the situation typical in many industries.

The presented use case relevant to the combination of expert knowledge with AI has shown that we can save resources by properly setting the reliability test variables. The approach proposed is mathematically complex and computationally intensive. Then, its implementation asks for a team of experts with various skills, always able to tackle issues at the forefront of research and innovation. Specific, further research work is needed to generalize the applicability

of the EIG-based Bayesian optimization framework to more complex decision problems in reliability engineering, such as considering sequential semi-Markov decision settings (e.g. due to the use of Weibull distribution) including censoring.

The presented causal modeling use case opens a new perspective on how to leverage the different information pieces relevant to maintenance, avoiding the ineffective strategies and waste of resources that can emerge from misinterpreting correlation as causation. This paves the way to further research work focused on developing causal modeling techniques for I5.0 and for reliability and maintenance issues, in particular. The proposed causal modeling application has also shown that it is extremely important that the model is rigorously defined and, then, that without a complete view of the context and the hypothesis that one is trying to prove, the results can become unreliable and contain biases that make them misleading. This highlights, once again, the need for a multidisciplinary team to develop successful applications.

Based on these findings, we think that the urgent measures that can be taken to help the industry adapt to the emerging context are i) investments in building multidisciplinary teams that can develop AI solutions addressing future, domain-specific challenges; ii) investments in R&D, to ensure that the AI solutions development teams can always handle the latest AI innovations; iii) evangelization of AI, to increase the number of AI final users, avoid misunderstanding on AI solutions final results and ensure that the AI solutions are given the right economic values; iv) throw caution to the wind, considering that the risks of an AI solution failing may be much lower than those of not having tried to develop it, i.e. losing the ability to navigate the current technological context and, therefore, not being ready to handle the challenges of the future industry.

References

[1] Xu X, Lu Y, Vogel-Heuser B and Wang L 2021 Industry 4.0 and industry 5.0-inception, conception and perception *J. Manuf. Syst.* **61** 530–5

[2] Ghobakhloo M, Iranmanesh M, Tseng M-L, Grybauskas A, Stefanini A and Amran A 2023 Behind the definition of industry 5.0: a systematic review of technologies, principles, components and values *J. Ind. Prod. Eng.* **40** 432–47

[3] Renda A European Commission, Directorate-General Research and Innovation 2021 *Industry 5.0, a Transformative Vision for Europe - Governing Systemic Transformations Towards a Sustainable Industry* (Publications Office of the European Union)

[4] Renda A 2020 *Enabling Technologies for Industry 5.0 - Results of a Workshop With Europe's Technology Leaders* (Publications Office)

[5] Breque M, De Nul L and Petridis A European Commission, Directorate-General Research&Innovation 2021 *Industry 5.0 - Towards a Sustainable, Human-Centric and Resilient European Industry* (Publications Office of the European Union)

[6] Shneiderman B 2020 Human-centered artificial intelligence: Reliable, safe & trustworthy (arXiv:2002.04087)

[7] Zio E and Guarnieri F 2024 Industry 5.0: Do risk assessment and risk management need to update? and if yes, how? *Proc. Inst. Mech. Eng.* O **OnLineFirst**

[8] Tamascelli N, Campari A, Parhizkar T and Paltrinieri N 2024 Artificial intelligence for safety and reliability: a descriptive, bibliometric and interpretative review on machine learning *J. Loss Preven. Process Ind.* **90** 105343

[9] Payette M and Abdul-Nour G 2023 Machine learning applications for reliability engineering: a review *Sustainability* **15** 7

[10] Ucar A, Karakose M and Kirimca N 2024 Artificial intelligence for predictive maintenance applications: Key components, trustworthiness and future trends *Appl. Sci.* **14** 2

[11] (Available at: https://en.wikipedia.org/wiki/TensorFlow)

[12] (Available at: https://en.wikipedia.org/wiki/PyTorch)

[13] Härlin T, Rova G B, Singla A, Sokolov O and Sukharevsky A 2023 *Exploring Opportunities in the Generative AI Value Chain* (McKinsey Digital)

[14] Information Technology Industry Council 2023 Understanding foundation models & the AI value chain: ITI's comprehensive policy guide (available at: www.itic.org/documents/artificialintelligence/ITI_AIPolicyPrinciples_080323.pdf)

[15] Bommasani R et al 2022 On the opportunities and risks of foundation models (arXiv:2108.07258)

[16] Taori R, Gulrajani I, Zhang T, Dubois Y, Li X, Guestrin C, Liang P, and Hashimoto T B 2024 "Alpaca: a strong, replicable instruction-following model, (available at: https://crfm.stanford.edu/2023/03/13/alpaca.html)

[17] Min B, Ross H, Sulem E, Veyseh A P B, Nguyen T H, Sainz O, Agirre E, Heintz I and Roth D 2023 Recent advances in natural language processing via large pre-trained language models: A survey *ACM Comput. Surv.* **56** 1–40

[18] OpenAI and Team 2022 Chatgpt: Optimizing language models for dialogue (available at: https://openai.com/blog/chatgpt)

[19] Chiang W-L *et al* 2023 Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality

[20] Wiggers K 2020 Openai's massive gpt-3 model is impressive, but size isn't everything (available at: https://venturebeat.com/ai/ai-machine-learning-openai-gpt-3-size-isnt-everything/)

[21] OpenAI and Team 2024 Gpt-4 technical report (arXiv:2303.08774)

[22] Bai Y *et al* 2022 Training a helpful and harmless assistant with reinforcement learning from human feedback (arXiv:2204.05862)

[23] Devlin J, Chang M-W, Lee K and Toutanova K 2018 Bert: Pre-training of deep bidirectional transformers for language understanding (arXiv:1810.04805)

[24] Tay Y, Dehghani M, Bahri D and Metzler D 2022 Efficient transformers: a survey (arXiv:2009.06732)

[25] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I 2017 Attention is all you need *Advances in Neural Information Processing Systems* vol 30

[26] Lewis P *et al* 2021 Retrieval-augmented generation for knowledge-intensive nlp tasks (arXiv:2005.11401)

[27] Fan W, Ding Y, Ning L, Wang S, Li H, Yin D, Chua T-S, and Li Q 2024 A survey on rag meeting llms: towards retrieval-augmented large language models (arXiv:2405.06211)

[28] Radanliev P, De Roure D, Maple C and Santos O 2022 Forecasts on future evolution of artificial intelligence and intelligent systems *IEEE Access* **10** 45280–8

[29] Klinger J, Mateos-Garcia J and Stathoulopoulos K 2022 A narrowing of AI research? (arXiv:2009.10385)

[30] (Available at: www.economist.com/interactive/briefing/2022/06/11/huge-foundation-models-are-turbo-charging-ai-progress)

048

[31] OECD 2023 *Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research* (OECD Publishing)
[32] (Available at: https://auto.gluon.ai/stable/index.html)
[33] (Available at: https://github.com/IBM/lale)
[34] (Available at: https://learn.microsoft.com/it-it/azure/machine-learning/concept-automated-ml?view = azureml-api-2)
[35] Pearl J 2009 *Causality: Models, Reasoning and Inference* 2nd edn (Cambridge University Press)
[36] Baltrusaitis T, Ahuja C and Morency L-P 2017 Multimodal machine learning: a survey and taxonomy *Trans. Pattern Anal. Mach. Intell.* **41** 423–43
[37] Karpatne A, Atluri G, Faghmous J H, Steinbach M, Banerjee A, Ganguly A, Shekhar S, Samatova N and Kumar V 2017 Theory-guided data science: A new paradigm for scientific discovery from data *IEEE Trans. Knowl. Data Eng.* **29** 2318–31
[38] Karpatne A, Jia X and Kumar V 2024 Knowledge-guided machine learning: current trends and future prospects (arXiv:2403.15989)
[39] Maddikunta P K R, Pham Q-V, Deepa P B N, Dev K, Gadekallu T R, Ruby R and Liyanage M 2022 Industry 5.0: a survey on enabling technologies and potential applications *J. Ind. Inf. Integr.* **26** 100257
[40] Bashir N, Donti P, Cuff J, Sroka S, Ilic M, Sze V, Delimitrou C and Olivetti E 2024 The climate and sustainability implications of generative AI *An MIT Exploration of Generative AI*
[41] Escobar L A and Meeker W Q 2006 A review of accelerated test models *Stat. Sci.* **21** 552–577
[42] Foster A, Jankowiak M, Bingham E, Horsfall P, Teh Y W, Rainforth T and Goodman N 2019 Variational Bayesian optimal experimental design *Advances in Neural Information Processing Systems* vol *32* (Curran Associates, Inc) pp 14036–47
[43] Foster A, Jankowiak M, O'Meara M, Teh Y W and Rainforth T 2020 A unified stochastic gradient approach to designing bayesian-optimal experiments (arXiv:1911.00294)
[44] Foster A, Ivanova D R, Malik I and Rainforth T 2021 Deep adaptive design: Amortizing sequential bayesian experimental design (arXiv:2103.02438)
[45] Foster A, Pukdee R and Rainforth T 2021 Improving transformation invariance in contrastive representation learning *Int. Conf. on Learning Representations*
[46] Lindley D V 1956 On a measure of the information provided by an experiment *Ann. Math. Stat.* **27** 986–1005
[47] Murphy K P 2022 *Probabilistic Machine Learning: An Introduction* (MIT Press)
[48] Murphy K P 2023 *Probabilistic Machine Learning: Advanced Topics* (MIT Press)
[49] Pearl J 2021 Causal and counterfactual inference *The Handbook of Rationality* p 427
[50] Peters J, Janzing D and Schlkopf B 2017 *Elements of Causal Inference: Foundations and Learning Algorithms* (The MIT Press)
[51] Pearl J and Mackenzie D 2018 *The Book of Why* (Basic Books)
[52] Bareinboim E, Correa J D, Ibeling D and Icard T 2022 *On Pearl's Hierarchy and the Foundations of Causal Inference* 1st edn (Association for Computing Machinery) pp 507–56
[53] Badman A 2024 Risk and the future of ai: algorithmic bias, data colonialism, and marginalization (available at: www.ibm.com/think/insights/ai-risk-management)
[54] Arora A, Barrett M, Lee E, Oborn E and Prince K 2023 Risk and the future of ai: Algorithmic bias, data colonialism and marginalization *Inf. Organ.* **33** 100478
[55] Radanliev P, De Roure D, Maple C and Ani U 2022 Super-forecasting the 'technological singularity' risks from artificial intelligence *Evol. Syst.* **13** 747–57
[56] AI-NIST 2024 Artificial intelligence risk management framework: generative artificial intelligence profile (available at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf)
[57] (Available at: https://ai-watch.ec.europa.eu/countries/italy/italy-ai-strategy-report_en)

Michele Compare is the CTO at Aramix, Milano, Italy, and the principal investigator of the projects on reliability, availability and maintenance analysis, optimization and decision making. He received the M.Sc. degree in mechanical engineering cum laude from University of Naples Federico II, in 2003, the PhD in nuclear engineering cum laude from Politecnico di Milano, in 2011. He was a research assistant at Politecnico di Milano. He worked as RAMS engineer and risk manager.

Enrico Zio received the M.Sc. degree in nuclear engineering from the Politecnico di Milano, in 1991, the M.Sc. degree in mechanical engineering from UCLA, in 1995, the Ph.D. degree in nuclear engineering from the Politecnico di Milano, in 1996, and the Ph.D. degree in probabilistic risk assessment from MIT, in 1998. He is currently a Full Professor with the Centre for Research on Risk and Crises (CRC), Ecole de Mines, ParisTech, PSL University, France, a Full Professor and the President of the Alumni Association, Politecnico di Milano, Italy. He has been or is distinguished guest professor at Tsinghua University, Beijing, China, adjunct professor at University of Stavanger and University of Tromso, Norway, City University of Hong Kong, Beihang University, Harbin Engineering University and Wuhan University, China, Kyung-Hee University, Korea. He is Co-Director of the Center for REliability and Safety of Critical Infrastructures (CRESCI) and the sino-french laboratory of Risk Science and Engineering (RISE), at Beihang University, Beijing, China. He is distinguished Research Fellow of the Institute of Nuclear Energy Safety Technology, in Hefei, China. He is member and vice-president of the Board of Directors of Fondazione Politecnico di Milano. He is IEEE and Sigma Xi Distinguished Lecturer. In 2020, he has been awarded the prestigious Humboldt Research Award.