Hybrid Intelligence in Next Generation Manufacturing: An Outlook on New Forms of Collaboration Between Human and Algorithmic Decision-Makers in the Factory of the Future

Frank T. Piller, Verena Nitsch, and Wil van der Aalst

Abstract The text discusses the concept of hybrid intelligence, which is a form of collaboration between machines and humans. It describes how this concept can be used in manufacturing to help improve productivity. The text also discusses how this concept can be used to help humans learn from machines. There is a debate in the intelligence community about the role of humans vs. machines. Machine intelligence can do some things better than humans, such as processing large amounts of data, but is not good at tasks that require common sense or empathy. Augmented intelligence emphasizes the assistive role of machine intelligence, while hybrid intelligence posits that humans and machines are part of a common loop, where they adapt to and collaborate with each other. The text discusses the implications of increasing machine involvement in organizational decision-making, specifically mentioning two challenges: negative effects on human behavior and flaws in machine decision-making. It argues that, in order for machine intelligence to improve decision-making processes, humans and machines must collaborate. The chapter argues that hybrid intelligence is the most likely scenario for decision-making in the future factory. The chapter discusses the advantages of this approach and how it can be used to improve quality control in a production system. The transformer-based

F. T. Piller (✉)
Institute for Technology and Innovation Management, RWTH Aachen University, Aachen, Germany
e-mail: piller@time.rwth-aachen.de

V. Nitsch
Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Aachen, Germany

Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE, Wachtberg, Germany
e-mail: v.nitsch@iaw.rwth-aachen.de

W. van der Aalst
Chair of Process and Data Science, RWTH Aachen University, Aachen, Germany
e-mail: wvdaalst@pads.rwth-aachen.de

F. T. Piller et al. (eds.), *Forecasting Next Generation Manufacturing*, Contributions to Management Science, https://doi.org/10.1007/978-3-031-07734-0_10

language model called GPT-3 can be used to generate summaries of text. This task is difficult for machines because they have to understand sentiment and meaning in textual data. The model is also a "few-shot learner," which means that it is able to generate a text based on a limited amount of examples. Transformer-based language models are beneficial because they are able to take the context of the processed words into consideration. This allows for a more nuanced understanding of related words and concepts within a given text.

[Abstract generated by machine intelligence with GPT-3. No human intelligence applied.]

1 From Human-Computer Interaction to Human-Machine Collaboration

As summarized in the previous chapter "Future Scenarios and the Most Probable Future for Next Generation Manufacturing", the most probable scenario resulting from our Delphi study on the future of digitalization in manufacturing predicts two fundamental changes until 2030 that will be enabled by the scaled deployment of digital shadows connecting data, products, and equipment across organizational boundaries: first, a shift from the current focus on operational efficiency to a broader set of economic, ecological, and social sustainability objectives driving future manufacturing strategies and second, an anthropocentric perspective on production where machines learn from humans and humans from machines in a much more collaborative form as compared to the status quo today.

In this final chapter of our book, we build on the second development. It corresponds to the paradigm shift from a technology-centered toward a human-centered digitalization and work design, consistently reconsidering the role of humans in the factory of the future (Mütze-Niewöhner et al., 2022; Hirsch-Kreinsen & Ittermann, 2021). Chapters "Organization Routines in Next Generation Manufacturing" and "Capability Configuration in Next Generation Manufacturing" already discussed these developments in larger detail. Human-centered digitalization and work design are also a central element of our understanding of an "Industry 4. U," as introduced in the first chapter of this book, describing the next evolution of Industry 4.0—centered on people *and* planet.

Human-centered digitalization starts with using technology to support humans at work in an individually customized manner by taking individual capabilities, habits, and preferences into account. Nevertheless, it also has a profound impact on how decisions are made in an organizational context, enabled by new forms of collaboration between humans and machines (machine intelligence). Delphi Projection P8 proposed the rise of a "hybrid intelligence," suggesting that in 2030, "strategic production decisions will be executed in close interaction between humans and AI-based algorithms." Our expert panel demonstrated consensus and a high probability that this projection will be realized within the next decade.

In this chapter, we explore the concept of hybrid intelligence in larger detail. While there are more questions than answers and we are just at the beginning to investigate this concept, early examples are already here. We used a specific use case of (a weak) hybrid intelligence to write this book: a (transformer) language model helped us to compose the abstracts and summaries of this book. While probably just a simple form of hybrid intelligence, it still provides a good illustration of a new form of collaboration between machines and us. We will discuss this specific application and its technical background in larger detail toward the end of this chapter. Before, we outline our understanding and definition of hybrid intelligence and the open research questions it poses with regard to the future organization of work. In this context, we present a specific scenario of using hybrid intelligence for learning and continuous improvement for Next Generation Manufacturing.

2 Hybrid Intelligence: Concept and Definition

There used to be a clear separation between tasks done by machines and tasks done by people (van der Aalst, 2021). Machine intelligence, i.e., mixtures of artificial intelligence (AI) and machine learning (ML), can deal amazingly well with unstructured data (text, images, and video) as long as there are enough training data. In the corporate context, the use of machine intelligence attempts to make structures and processes more efficient. Applications in speech recognition (e.g., Alexa and Siri), image recognition, automated translation, autonomous driving, and medical diagnosis have blurred the classical divide between human tasks and machine tasks. However, while machine intelligence works well for such clearly defined tasks, it is not foreseeable that it will become capable of fully mapping complex business problems in organizational contexts (Dellermann et al., 2019) or solving multiple tasks simultaneously (Raj & Seamans, 2019). Although current AI and ML technologies outperform humans in many areas, tasks requiring common sense, contextual knowledge, creativity, adaptivity, or empathy are still best performed by human intelligence. Machine intelligence, on the contrary, is about data and algorithms and can be characterized by terms such as fast, efficient, cheap, scalable, and consistent.

Taken together, Dellermann et al. (2019) define *hybrid intelligence* as:

> the ability to achieve complex goals by combining human and artificial intelligence, thereby reaching superior results to those each of them could have accomplished separately, and continuously improve by learning from each other.

Following this definition, hybrid intelligence hence blends human intelligence and machine intelligence to combine the best of both worlds. As things stand today, it is the most likely deployment scenario of machine intelligence in the corporate context over the next few decades. Hybrid intelligence aims to leverage the complementary strengths of human and machine intelligence in such a way that better overall performance can be achieved than when machines or humans are used alone (Dellermann et al., 2019; Kamar, 2016). Even in often-cited application scenarios that use AI-based algorithms for decision preparation or outsource decision-making to AI (e.g., laboratory data interpretation, human resources, claims processing), human actors invariably play a central role (Shrestha et al., 2019).

A closely related term, *augmented intelligence*, emphasizes the assistive role of machine intelligence (especially ML), when deep neural nets and other data-driven techniques enhance human intelligence rather than replace it. In this understanding, AI and ML are shifting human intelligence on a higher level, just like telescopes are there to enhance human vision. The term is widely used especially in the literature on computational medicine for algorithms supporting humans in medical diagnosis and research. Long and Ehrenfeld (2020) proposed such an augmentation scenario impressively for the case of reacting to the Corona pandemic (in a paper published at a time when the general public hasn't realized yet that there was a pandemic), forecasting a coordinated research endeavor to fight the spread of the disease that would have been not possible without strong ML capacities supporting the research teams. Reality proofed their predictions right.

However, in the understanding of augmented intelligence, there still is a sequential process in the division of labor between humans and machines: Machines process large amounts of data, search for patterns, and make predictions, but basically support humans, who drive the process, and execute the results of the AI. Our understanding of hybrid intelligences goes further, regarding human and machine intelligence as two elements of a common loop. In doing so, we follow the definition by Dellermann et al. (2019), as presented above, or Zheng et al. (2017), who describe a "human-in-the-loop hybrid-augmented intelligence" system, where humans are always part of the system. In this system, humans first influence the outcome (of a machine intelligence) in such a way that they provide further judgment if a low confident result is given by the algorithm. But the collaboration goes further. The idea is to "realize a close coupling between the analysis-response advanced cognitive mechanisms in fuzzy and uncertain problems and the intelligent systems of a machine" (Zheng et al., 2017: 154). Hence, human and machine intelligence adapt to and collaborate with each other, forming a two-way information exchange and control (a similar understanding has been outlined by Pan (2016) in his conceptualization of an "Artificial Intelligence 2.0"). This is why we prefer to use the term *hybrid* (and not augmented) intelligence.[1]

A good illustration of this collaboration between human and machine intelligences provides AlphaGo, a Go-playing computer developed by DeepMind Technologies (a firm belonging to Alphabet Inc., the mother company of Google). Commonly seen as a breakthrough in machine intelligence, AlphaGo defeated the

[1] Zheng et al. (2017) also describe a second concept of human and machine collaboration: "cognitive computing-based hybrid-augmented intelligence." While out of the scope of this chapter, it is worth mentioning. Cognitive computing-based hybrid-augmented intelligence refers to a machine that "mimics the function of the human brain and improves computer's capabilities of perception, reasoning, and decision-making. In that sense, [it] is a new framework of computing with the goal of more accurate models of how the human brain/mind senses, reasons, and responds to stimulus, especially how to build causal models, intuitive reasoning models, and associative memories in an intelligent system" (Zheng et al., 2017: 154).

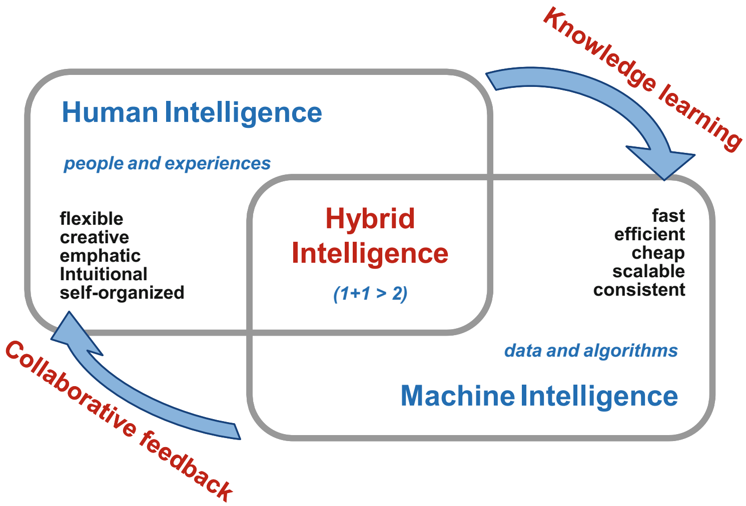

Fig. 1 Hybrid intelligence (HI) aims to combine the best of human intelligence and machine intelligence [Source: Building on van der Aalst (2021) and Zheng et al. (2017)]

best-ranked Go player Ke Jie in 2017. The more powerful AlphaGo Zero learned by just playing games against itself but was able to defeat any human player by the end of 2017. However, this has not been the end of the story (van der Aalst, 2021). The interplay between human intelligence and machine intelligence led to new insights. AlphaGo showed human players new strategies for playing Go, as some of the world's leading Go players acknowledged [as recorded in Baker and Hui (2017)]. Shi Yue said "AlphaGo's game transformed the industry of Go and its players. The way AlphaGo showed its level was far above our expectations and brought many new elements to the game." Zhou Ruiyang said "I believe players more or less have all been affected by Professor Alpha. AlphaGo's play makes us feel freer and no move is impossible to play anymore. Now everyone is trying to play in a style that has not been tried before." At the same time, the new strategies explored by the human players inform the machine algorithm. Humans can learn from machines, and machines from humans: "We look forward with great excitement to AlphaGo and human professionals striving together to discover the true nature of Go," Baker and Hui (2017) conclude a review of the innovations to the gameplay of Go, resulting from the collaboration of human players and the AlphaGo machine.

Hybrid intelligence aims to combine the best of both worlds, as illustrated in Fig. 1. The recent developments in AI and ML have extended the reach of software and hardware automation (robots). Once a robot is able to perform a repetitive task at a similar level of quality, it is often also more reliable and cost-effective. However, humans still have unique capabilities. For example, we have the ability to transfer experiences from one problem domain to another. As van der Aalst et al. (2021) argue, AI/ML cannot deal with disruptions. The Corona pandemic or events of severe weather like the flooding in Germany in July 2021 have shown that when there is a sudden dramatic change, predictive models fail, no matter how much data was there before. Especially at the beginning of the Corona pandemic, the established algorithms predicting demand in supply chains failed because of the unforeseen demand for certain products (e.g., pasta and toilet paper) combined with simultaneous restrictions for travel, work, and business. In such a situation, machine intelligence needs to be complemented by human intelligence.

But also in non-catastrophic events, humans need to remain in the loop. The idea of hybrid intelligence is not just to use humans when machine intelligence fails due to disruptions. The allocation of machine intelligence in decision-making processes often leads to more efficient, but sometimes also to unreflective or non-transparent, solutions with unintended biases. This, in turn, leads to a rejection of the AI contribution (acceptance) and thus hinders the exploitation of its potentials. Consider situations that need empathy, creativity, or ethics (van der Aalst, 2020). Decisions in these situations will also demand human contributions and cannot entirely be executed by a machine. Machine intelligence and human intelligence will complement each other. Understanding these factors as well as the mechanisms of interaction between humans and machine intelligence is a domain that opens a wide demand for further research. We will explore these dynamics in larger detail in the following section.

3 New Rules for Task Allocation: Division of Labor Revisited

The rise of hybrid intelligence asks us to reconsider one of the most fundamental of all economic and ergonomic questions: the division of labor and task allocation in an organization and individual work systems. While the development of machine intelligence is a field of computer science (decision routines and data structures) and research on corresponding technical applications of AI is primarily located in the engineering sciences, the implementation of hybrid intelligence is an economic (management) phenomenon (Bailey & Barley, 2020; von Krogh, 2018). It asks the question how to efficiently design decision-making in an organization.

Since the days of Frederick Taylor and Henry Ford, the idea of the ideal human-machine task division has evolved considerably from an industrial engineering and ergonomic perspective. Machine intelligence has the potential to be more than a tool, as it can also take on the role of a work partner or even a supervisor, as suggested in the debate of algorithmic management (Lee et al., 2015). In a work system, humans and AI need not oppose each other, but can complement each other as a team. Still, today, humans are only used for monitoring systems automated by machine intelligence. These humans are either under-challenged or fatigued, which significantly prompts errors. Other humans, who already are heavily burdened by their own subtasks, get overwhelmed by the need to make additional decisions as to when AI support should be utilized. Hence, to effectively support and relieve humans, machine intelligence should therefore work largely independently and recognize when support is necessary and desired. Furthermore, a dynamic division of tasks between humans and machines could adapt to varying situations, tasks, and user states, avoiding states of cognitive overload and underload. As a basis for such adaptive support, data providing information about the states of the individual components of a work system, like the involved human(s), equipment, the environment, as well as task and organizational goals, are needed and can be provided in the future in the form of digital shadows.

When the extent of decision support by machine intelligence is reaching intensity levels that seemed impossible in the past, research is needed how tasks can be allocated in the continuum between machine intelligence and human intelligence. Prior research in this domain rather described the challenge based on a few case studies (e.g., Iansiti & Lakhani, 2020; De Cremer, 2020) or exploratory surveys (Berditchevskaia & Baeck, 2020) and rather focused on the practical implementation of decision processes with machine intelligence, but neither examine their organizational impact nor do they follow the understanding of a hybrid intelligence, as discussed before.

We propose to structure such a research endeavor into two dimensions:

1. What is the (optimal) degree of integration of machine intelligence into organizational decision processes, and what are the tasks remaining for humans and the tasks where a human-machine collaboration is the preferred solution?
2. What is the quality of decisions made by the use of machine intelligence—not just when compared to the factual quality of the decision for a given task (if benchmarked against human decisions) but also when taking factors of organizational acceptance and adoption of the machine decision into account?

3.1 Degree of Machine Intelligence Integration into Organizational Decision Processes

To analyze the degree to which machine intelligence is involved in organizational decision-making, the established logic of the automation pyramid in engineering provides a good framework (Endsley, 1987). Consider the different cases shown in Fig. 2. The two extremes are the established situations of human and machine intelligence. But as the picture shows, there is a scope of hybrid situations [(b) to (d) in Fig. 2]. Here, to varying degrees of intensity, human and machine intelligence interact, each with particular strengths (and weaknesses) and major differences in capabilities and behaviors, in ways that did not exist in earlier human-human interactions (Berditchevskaia & Baeck, 2020; Groensund & Aanestad, 2020). In a narrow understanding of our definition of hybrid intelligence, only Case (d) addresses the intended collaboration between human and machine intelligence; Cases (b) and (c) are rather situations of "augmented intelligence." However, the

	Decision preparation	Decision	Decision scenario and assessment
Human Intelligence	***Human***		***(a) Human makes the decision alone.*** • Ability to judge and interpret ("common sense") also with regard to ethical standards; identification of objectives; creativity, flexibility; emotional intelligence • Possibly technically inadequate, slow; characterized by human bias in decision-making.
Hybrid intelligence	***AI & ML (machine)***	***Human***	***(b) Information is prepared by machine intelligence, but the actual decisions are made by humans*** *("augmented intelligence")* • Human judgment as a controlling factor when using ability for, e.g., pattern recognition and speed of AI&ML • Acceptance of machine prediction / prescription harmed by interpretation, norm violations, distribution effects, non-transparency
	Human	***AI & ML (machine)***	***(c) Humans provide machines with information, but autonomous decision-making is done by machine intelligence*** • Initiation by humans (control of the process), by, e.g. providing pre-selected or pre-labeled data • Solution may contain human bias; not based on sufficient data
	Human-AI (human-machine) Collaboration		***(d) Machine and human intelligence are integrated and collaborate during the decision process and task execution*** • Iterative learning between human and machine based on machine-enabled pattern recognition and simulation, as well as human feedback to the machine. • Machine decision patterns shape human decision strategies in the long term
Machine Intelligence	***AI & ML (machine)***		***(e) Decision is fully delegated to an autonomous machine*** • Speed, scalability, processing of large amounts of data. Objectivity (i.e. ex ante free from human bias). • Not always technically feasible; erroneous, especially biased decisions with distorted or incomplete data.

Fig. 2 Different situations of combining human and machine intelligence

borders between these areas are fuzzy and constantly moving, as we will illustrate with a simple example at the end of this chapter.

All situations of hybrid intelligence have immediate consequences for the behavior of individuals and thus for the resulting (quality of the) decisions and their implementation. In the longer term, they will also result in indirect effects, when people's experiences with machine intelligence influence their subsequent behavior in other situations (e.g., always expecting that there is a machine intelligence at hand to support a human task). Also, undesirable path dependencies may arise, such as a loss of knowledge or skills (Lebovitz et al., 2022), as experienced by the use of GPS-based navigation systems, which deterred the ability of many humans to navigate without machine support.

Hence, a critical question is when the potential benefits of allocating decision-making tasks to machine intelligence (increasing the efficiency and effectiveness of the decision-making process) are (over)compensated by new costs and challenges. These costs include both the efforts for developing and implementing the algorithms and the cost of adapting an organizational design to the new situation. Also, indirect costs in the form of negative effects on human behavior must be considered, e.g., costs resulting from acceptance problems. Acceptance here addresses both the individual level, i.e., humans who must share decision power with machine intelligence and collaborate with it, and the societal level of acceptance by stakeholder groups such as trade associations, unions, or regulatory institutions.

3.2 Consequences for Decision Quality

For certain, well-defined decision situations and tasks, machine intelligence provides without doubt better results, i.e., adds real value (without obvious violation of norms and other constraints). However, also in these situations, a remaining challenge is often the black-box nature of the solution (Shrestha et al., 2019). In computer science, approaches are therefore being developed to make AI more comprehensible (Rai, 2020), so that people are more likely to accept and implement the solution provided by the machine intelligence (when completely autonomous task performance is not possible/desirable). Scenarios of using hybrid intelligence are obvious in these decision-making situations.

In other situations, however, it is not certain whether machine intelligence can provide a suitable and better solution. This may be because (1) relevant norms to the decision are not observed by the machine and/or (2) the technical solution is "flawed," because the underlying data basis is insufficient or the modeling has not adequately captured the problem or cannot capture it due to unknown causal relationships. An example of such flawed decisions can be found in recruitment. When past career paths and performance patterns are used as the basis for future hiring, women tend to be left out of the equation (Cowgill & Tucker, 2020). This results in a conflict with the social norm of increasing diversity. The reasons behind these flawed decisions can be insufficient amounts of data or discriminatory patterns contained therein, but also an ill-defined notion of recruitment performance. However, once such a problem has been understood, humans together with machine intelligence can improve automating these decisions in the mid-term.

We believe that this situation also reflects the reality in most manufacturing companies today (Agrawal et al., 2019; Raj & Seamans, 2019). Machine intelligence is used but requires collaboration with human decision-makers to result in an optimal solution. Hence, an important question is how humans could check the quality of prescriptions provided by a machine, considering a potential violation of norms or possible "errors," before implementing the solution in a corrected manner, a procedure that Groensund and Aanestad (2020) called "augmenting the algorithm." As we will argue in the next section, real-time simulation models enabled by digital twins and shadows allow exactly such an ex ante validation. At the same time, structuring a machine intelligence solely according to human thought patterns (or those that humans can understand) is not sufficient either, as it may model the problem task inadequately or follow violations of norms by human decision-makers. This is exactly where the vision of a hybrid intelligence comes into place. Once the issues outlined before are recognized and understood, either an autonomous decision process by machine intelligence could be improved, or the decision could be structured in such a way that humans stay in the loop, taking social norms or intended consequences into account. Equally, however, humans also improve their own decision-making processes, when, for example, a machine intelligence suggests previously unknown initial solutions or uncovers distorted decision-making patterns of humans in the past. The loop is closing.

4 Hybrid Intelligence in Next Generation Manufacturing

While we believe that hybrid intelligence will strongly influence all kinds of decisions and task execution in an organization, we want to demonstrate such a scenario for Next Generation Manufacturing, as central to this book. As introduced in chapter "How Digital Shadows, New Forms of Human-Machine Collaboration, and Data-Driven Business Models Are Driving the Future of Industry 4.0", the context of this work is the interdisciplinary research cluster *Internet of Production (IoP)* at *RWTH Aachen University* (iop.rwth-aachen.de), enabling a new level of cross-domain collaboration along the entire product life cycles from engineering over operations toward the usage stage (Brecher et al., 2016). The IoP pursues a vision called the World Wide Lab (WWL), in which processes, factories, entire companies, and the managers and workers constituting these organizations can learn from each other by sharing experiences and knowledge (Brauner et al., 2022). Corresponding to the relationship of the Internet and the World Wide Web (WWW), the WWL aims to be a network of multisite labs in which models and data from experiments, manufacturing, and usage are made accessible across company borders to gain additional knowledge. A main driver of the WWL is *digital shadows*, i.e., purpose-driven, aggregated, multi-perspective, and persistent data sets from production, development, or usage (Liebenberg & Jarke, 2020). Digital shadows are a specification of the broader idea of digital twins (for more details, refer to chapter "How Digital Shadows, New Forms of Human-Machine Collaboration, and Data-Driven Business Models Are Driving the Future of Industry 4.0"). The cross-domain exchange of digital shadows in the form of *data spaces* can make data more valuable, opening the present data silos in different companies—a core enabler of better machine intelligence.

In our understanding of the Internet of Production, digital shadows are the "units of data" shared among organizations. They connect data, products, and industrial assets within and across organizations and are the foundations for data-driven planning and decisions within an organization (factory) and in-between organization (supply chains, value chains) by using real-time and historical data to simulate predicted futures. In this loop, hybrid intelligence plays a central role. Figure 3

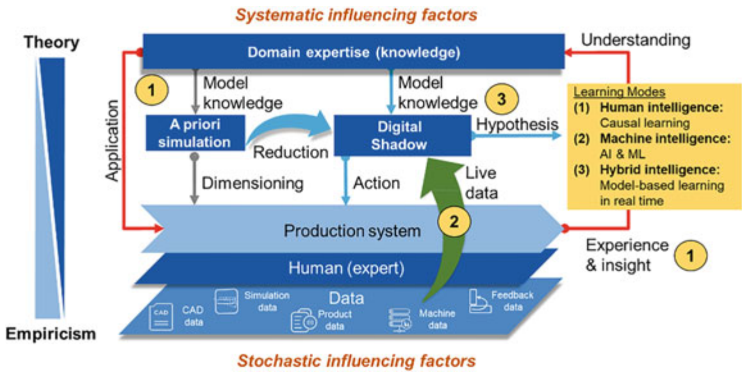

Fig. 3 Three models of learning in manufacturing: (1) causal learning, (2) machine learning, and (3) model-based learning based on digital shadows [building on Brecher et al. (2017)]

outlines such a *hybrid decision-making* combining machine intelligence and human expertise in a collaborative form. The figure shows three different modes of learning (understood here generically as any kind of decision-making in an existing manufacturing system to improve the system's operational efficiency, to cope with disturbance, or to increase the system's potential for strategic differentiation).

1. *Human intelligence: Causal learning* is the established way we learn. Building on *domain knowledge* acquired either by experience (learning on the job) or by formal education, humans have a unique capability to understand a complex system and utilize or improve it by trial-and-error learning. The experience curve effect is based on this learning mode, as are practices like Lean Six Sigma. Informed by their domain expertise, a team at a production station defines a problem area (an *application*), sets up *assumptions* (hypotheses) how to achieve an improvement, tests the assumptions via *experiments* to gain *insights*, and then either implements the solution (if the experiment was successful) or redefines the assumptions and conducts a new experiment. The development of the hypotheses is based on *theory*, often captured in models of the systematic influencing factors of the production system (like fluid or thermal dynamics) and uses the real production system as the test bed for empirical validation (*empiricism*). Such a causal learning process can be very powerful, but it is often slow and prone to the assumptions human draw and the hypotheses they set up.

   Conventional (digital) simulation models also belong to this learning mode. An *a priori simulation* uses ("theoretical") model knowledge to simulate (an extract of) the production system, so that specific behaviors (assumptions, scenarios) can be tested. These digital models can be used to reason about reality and answer what-if questions. However, digital simulation models are a reflection of reality that is created manually and functions in an offline manner, i.e., the model does not change when reality changes (van der Aalst et al., 2021). Hence, conventional simulation models are outdated when the production system goes into operations, as there are numerous *stochastic factors* influencing the system behavior, like temperature conditions, material characteristics, or the mood of the humans involved. All of these factors lead to a continuous change of the system (like an abrasion of a component, minimal modifications of a material, etc.). Conventionally, these changes are not captured in the simulation model, which is why experiments in the real system are required.

2. *Machine intelligence: Artificial intelligence and machine learning* came up as the new way to learn. Machine intelligence is data-driven and learns from data without explicitly being programmed. In manufacturing, cyber-physical systems provide these data in real time (in the form of digital twins and shadows) and store it in repositories (data spaces) where algorithms can find insights and interrelations between different data sets. Today's usage of machine learning for many tasks that before could only be done by humans can be attributed to progress in deep learning techniques, where *artificial neural networks (ANNs)* having multiple layers progressively extract higher-level features from the raw input (van der Aalst, 2021). For example, we can train an ANN to distinguish between pictures from a vision control system that show work pieces with adequate and others with insufficient quality. While training, the ANN updates the weights in the internal representation until the number of incorrectly classified pictures is minimized. Then the trained ANN is used to classify test data, i.e., unseen pictures of good and bad pieces that need to be classified correctly. Given enough training data, such an ANN may perform amazingly well in automating quality control, although it was never programmed to do so and has no explicit knowledge of what makes a good and a bad work piece.

   Beyond such automation scenarios enabled by machine intelligence, also higher-level learning can take place. When the quality data (from the automated vision control system) is matched with data from other workstations of this production system, algorithms can find patterns between two system elements, identifying also states in one production step that causes later whether a work piece is labeled as good or bad. This ability of finding patterns in huge data sets led some people to say that the future of learning in manufacturing is only pattern recognition in huge data spaces—no human input and no domain knowledge required. However, we believe—and were confirmed by the results of our Delphi study—that such a pure machine intelligence scenario is unlikely to cope with the complexity of a real production system.

3. *Hybrid intelligence: Model-based learning in real time* is our proposed scenario for learning in Next Generation Manufacturing. Without doubt, machine intelligence can perform repetitive operational tasks more efficiently than humans can. Machine learning algorithms also have an unmatched capability of finding patterns in large data sets. We propose that these insights generated by *machine intelligence* serve as a highly educated "hunch" for humans, who combine it with their domain expertise on a higher level. An important component of this approach is the availability of *digital shadows* as virtual, real-time digital counterparts of something that exists in the physical world (e.g., a production system, workstation, or work piece). The digital counterpart should help to make decisions in a better way, by not providing the real-time data from which a machine

intelligence can generate its insights, but also the test bed where ideas for improvement and optimization can be validated virtually.

Consider the quality example from Scenario (2). Let us assume that an algorithm provided an insight in the form of a prediction on the causes of a quality issue: "When the temperature in Station A dropped below a specific threshold, later quality errors occurred in Station E." With the availability of a digital shadow, and different to conventional digital simulation models as in Scenario (1), the model behind the digital shadow is automatically derived and changes when reality changes. The digital shadow can now be used to reason about reality and answer what-if questions. Hence, assumptions on how to improve the quality issue in our example can be tested virtually in the simulation models embedded in the digital shadow. This connects Scenarios (1) and (2). Based on their *intuition* and *domain expertise*, human decision-makers could make conclusions on how to improve the quality of the system, e.g., different approaches to control the temperature in Station E in a more stable way or approaches to counterbalance the temperature effect on to work pieces in later word stations. These assumptions about how to improve the system's quality, provided by *human intelligence* but augmented by insights generated by *machine intelligence*, could now be validated in the virtual shadow. The virtual experimentation allows testing of many more alternative scenarios for improvement. A machine intelligence could support this experimentation, e.g., by proposing different scenarios and predicting their outcomes.

In a further state, an *automated real-time feedback loop* can be established. The insights produced by the digital shadow could then either automatically trigger changes in the production system or become implemented manually by humans after interpreting the results (van der Aalst et al., 2021). Results of the digital shadow directly affect reality. For operational situations, autonomous learning and optimization is likely. For example, when the simulation model predicts a delay, the production process could be reconfigured automatically (similar to the re-routing algorithm in a navigation system when it is informed about an incident on the originally planned route). For more complex learning scenarios, like restructuring the manufacturing system or coping with disruptions, the advanced simulation model embedded in a digital shadow allows human decision-makers to evaluate all possible decisions in the virtual world without causing harm, waste, and costs in the real (physical) system. With cheaper and richer experimentation, the likelihood of finding a better solution increases.

We have to stress that this scenario is a picture of the future yet, especially when we apply it on the level of a larger system. In our research in the Internet of Production cluster at RWTH Aachen, our colleagues were able to demonstrate this approach on the component and work station level (Brecher et al., 2019; Xi et al., 2021). *Process mining* can serve as a concrete technology to facilitate the development of such a virtual shadow/twin of an entire system (van der Aalst, 2016). Using process discovery, so-called control-flow models can be derived. Aligning these models with event data, it is possible to add different perspectives (time, costs, resources, decisions, etc.). The resulting elaborate model can be simulated. Using process mining, it is relatively easy to create a digital shadow in terms of a frequently updated virtual replica of a physical object. However, it still is rather difficult to create a model that behaves like a real system, where multiple processes interact and compete for resources concurrently. To fulfill the vision of a digital shadow that automatically takes action, action-oriented process mining provides initial ideas (e.g., the *Celonis Execution Management System* can trigger corrective workflows using the *Integromat* integration platform). But despite these initial capabilities of process mining, it is fair to say that this scenario of hybrid intelligence is more a vision than a reality. We need to keep humans in the loop (Abdel-Karim et al., 2020) to cope with the complexities of an entire production system. This is why we regard hybrid intelligence as the most likely scenario for decision-making the factory of the future.

5 A Simple Application of Hybrid Intelligence in Publishing

We want to close this chapter by a simple use case of hybrid intelligence. When writing and producing this book, we recruited an AI as a member of our author team, tasking it with creating all abstracts of this book's chapters and writing the book's preface. This worked amazingly simple, providing us a real glimpse into a future where machines and humans collaborate intuitively.

The AI we used is a transformer-based language model. While quantitative data prevails in a production context, much knowledge is shared through natural language. By talking to a colleague, listening to a lecture, or reading a book, understanding language grants us access to a plethora of knowledge. Today, AI has reached a good level of language understanding, so that we can use such technologies to further share and create knowledge. This makes language models an especially interesting form of AI to use in knowledge-intensive work (Bouschery et al., 2022).

Transformer-based language models are a special kind of AI used for natural language processing (NLP), which Liddy (2018: 3346) defines as a range of "computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications." In general, natural language processing is not new to firms. It has been used, for example, in text analysis (text mining), like generating insights from maintenance or service reports. Prior models have typically been very task specific. Also in this field, a great deal of progress stems from advances in ANNs. Newer NLP technologies show the potential to take on multiple knowledge-related tasks and cannot just analyze existing text but also generate new one. A core example of these advanced NLP models is generative or transformer-based models. At its core, language modeling is the process of predicting the next word in a sequence based on its preceding characters or words. This field has seen continued progress over the past decades with a trend toward larger and more complex models rapidly increasing the models' capabilities—from the mere suggestion of related words to state-of-the-art models that can produce full newspaper articles indistinguishable from human-written text (Brown et al., 2020).

New transformer-based models contain attention mechanisms that allow for parallelization during the processing of inputs and thereby eliminate some of the main performance issues of recurrence-based models, leading to significantly faster models (Vaswani et al., 2017). Another big advantage of these types of models is their ability to take the context of the processed words into consideration, which allows for a far more nuanced understanding of related words and concepts within a given text and subsequently more complex applications. Today, most of the state-of-the-art language models are based on this transformer architecture and rely on large data sets only for pre-training purposes. Examples include Google's BERT model or OpenAI's line of Generative Pre-trained Transformers (GPT). The number of parameters used to generate the models' output has increased significantly over the last few years. For example, the original BERT model (Devlin et al., 2018) uses 340 million parameters in its largest instance. This pales in comparison to OpenAI's latest model, GPT-3. In just 3 years, the model size of the GPT line has grown by nearly 1600% from 110 million parameters in the original model over 1.5 billion parameters in its second iteration (Radford et al., 2019) to 175 billion parameters in GPT-3 (Brown et al., 2020). The next version is expected to have 100 trillion parameters. Because transformer-based language models' capabilities significantly improve with model size, the rapid increase in model sizes has dramatically increased the usefulness and applicability of such transformer-based language models.

While it is very cost-intensive to build and train large transformer-based language models in the first place, many of these models have been open-sourced and can be accessed very easily through web services, making them accessible to a broader audience. Also, commercial applications like GPT-3 are available in cloud-based applications via a standard Internet browser. Another big advantage of transformer-based language models is that users can generally interact with them simply through natural language. Companies like OpenAI provide access to their models through not only application programming interfaces (APIs) but also graphical user interfaces (GUIs), which significantly lower the barriers to entry.

For this book, we utilized OpenAi's GPT-3 [we refer to Bouschery et al. (2022) for a more detailed description of our approach]. To interact with the model, users have to provide some initial text input. This could either be a question, the beginning of a story that should be completed, some text that should be summarized, bullet points to turn into written text, etc. Based on this initial input and its knowledge learned during training, GPT-3 then generates a text that best fits the provided prompt by predicting the next word in the sequence based on the previous words in the prompt. GPT-3 is a so-called few-shot learner, which means that users are advised to provide the model with a few examples to show what kind of output they expect from the model. The initial prompt is therefore the main way of steering the model toward a desired output—*a perfect illustration of a hybrid intelligence*.

We hired the GPT-3 to become a member of our publishing team for a typical knowledge processing task: knowledge extraction (De Silva et al., 2018), i.e., making existing knowledge usable by extracting knowledge that might be coded explicitly or implicitly in a given knowledge base. Normally, extracting knowledge is rather labor-extensive and not easily scalable. However, transformer-based language models provide the opportunity to automate parts of such processes. The knowledge task we asked the GPT-3 to do was *text summarization*. All abstracts of this book (including the one of this chapter) were generated by the GPT-3. Extracting knowledge from text requires the identification, selection, and processing of large amounts of textual data, which can be very costly and time-consuming as well as overwhelming for humans. This might lead to unwanted consequences like bad decision-making, as humans tend to make simplifying assumptions in order to be able to process such large volumes of information (Simon, 1955). Transformer-based language models provide an adequate solution to such problems through their ability to summarize text efficiently while conveying the most important aspects and meaning of the original text. They are especially well suited for such tasks because of their context awareness and the ability to pay special attention to the most relevant pieces of information. This gives them a great advantage over earlier text summarization algorithms, which struggled to understand sentiment and meaning in textual data (Fan et al., 2006). We invite our readers to evaluate the results by themselves.

Furthermore, users can prompt state-of-the-art language models like GPT-3 to summarize a text in a very specific way. In addition to the abstracts provided in each chapter of this book, we also tasked GPT-3 to summarize the entire book as short as possible to test an extreme case of summarization. This is the result [the differences result from varying settings of the model, see Bouschery et al. (2022), for more details].

Prompt: "`GPT-3, summarize this book in a few sentences.`"

- The text discusses different projecting made by experts about the changes that will occur in production by 2030. These include the trend of increasing digital goods, the use of hybrid intelligence to supplement human decision-making processes, the establishment of full digital twins of each production worker and all of their operations, and the decentralization of supply chains.
- The text discusses various trends that are projected to occur in the manufacturing industry by 2030. These trends include the use of subscription models for production machines, the use of collaborative robots, the use of digital twins, and the decentralization of supply chains. Each of these trends is expected to have a significant impact on firms.
- The Delphi study found that in 2030, subscription models for production machines will be the new industry standard, that environmental sustainability will play a bigger role, and that AI-based decision-making systems will not be able to significantly increase production resilience.

Prompt: "`GPT-3, summarize this book in one sentence.`"

- The text discusses various projections for the year 2030 and their potential impacts on different aspects of society and the economy.

These examples show the ability of transformer-based language models to capture the essence of a text and reproduce a coherent, yet substantially summarized, version of the original text. We did not provide any further context and examples or used fine-tuning to produce the response. While this task might seem rather trivial to a human, it is rather complex. First, the model had to understand that there is a difference between the provided article and the task description. Specifically, the task descriptions are, in fact, exactly that, a description of what the model should do, and not, e.g., the beginning of a novel that should be continued. Then, it had to analyze, understand, and then generate a summary of the said abstract that was factually, semantically, and grammatically correct. All, without having specifically been trained to perform this task. Noteworthy is also that the model did not just shorten the provided text, but that it summarized the text in its own words. However, when looking closely at the generated texts, we instantly find expressions which we would write differently, where there would be a dedicated technical term to describe the subject more precisely for an expert audience, or where we also would emphasize an aspect we believe being most interesting for our target audience of academic peers (who the algorithm does not know at all).

Hence, we propose that transformer-based language models will specifically support knowledge-based practices in the form of a hybrid intelligence. Their ability to interact with different knowledge sources, to learn from them, and to transform knowledge allow these models to act as a knowledge broker that facilitates the sharing of knowledge between different stakeholders while also fostering the creation of new knowledge (Waardenburg et al., 2022). Human teams can employ these language models to access existing knowledge. Models that have been trained on large text corpora from the Internet have knowledge on a wide range of topics, which opens up the opportunity for teams to integrate knowledge that might lay outside their area of expertise. Given a prompt by a human, the AI can help to establish connections between concepts and ideas that might otherwise not have been obvious. Few-shot learning capabilities then allow for an easier interaction between the humans in a hybrid team and the AI. Humans have to provide a limited number of exemplary responses to a given task, so that the language model can generate a first adequate output. Humans then evaluate this output, indicating to the algorithm, for example, parts of the output they find especially interesting. The algorithm will then produce a next output, based on this feedback. In the true understanding of a hybrid intelligence, machines and humans are building upon each other's input and output.

In such a scenario, teams can integrate the AI in their existing processes, as if it would be a new colleague. The combination of domain expertise by human team members and knowledge provided by the AI provides the opportunity to greatly improve productivity of knowledge-based practices and produce outcomes that would not have been possible with just the skillset of one of the actors. Orchestrating and building such hybrid teams becomes a new important managerial task, and understanding when and how to allocate tasks to a machine intelligence (and which one) will be a key success factor of organizations in the future. Managers have to consider the distinct characteristics of human and non-human actors. While humans will play a major role in providing context, steering language models toward desired results, and embedding AI output in the larger picture, machine intelligence can speed up many tasks that require the handling of large amounts of text (or other data), understand patterns in data invisible to humans, and make connections between knowledge bases that might not be readily available to human team members.

While for more complex tasks like steering a production system such a scenario of hybrid intelligence is still not existing, the way of development seems clear. We hope that this chapter, but also the analysis of our Delphi study in the entire book, provides the reader plenty of ideas and food for thought about the future of industrial production and the elements of Next Generation Manufacturing.

Acknowledgment Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC-2023 Internet of Production—390621612.

References

Abdel-Karim, B. M., Pfeuffer, N., Rohde, G., & Hinz, O. (2020). How and what can humans learn from being in the loop? Invoking contradiction learning as a measure to make humans smarter. *Künstliche Intelligenz, 34*(2), 199–207. https://doi.org/ghqvr8

Agrawal, A., Gans, J., & Goldfarb, A. (2019). Artificial intelligence: The ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives, 33*(2), 31–50. https://doi.org/ggnh5t

Bailey, D., & Barley, S. R. (2020). Beyond design and use: How scholars should study intelligent technologies. *Information and Organization, 30*(2), 100286. https://doi.org/hm62

Baker, L., & Hui, F. (2017, April). *Innovations of AlphaGo*. Research blog by Deepmind. https://deepmind.com/blog/article/innovations-alphago

Berditchevskaia, A., & Baeck, P. (2020). *The future of minds and machines: How AI can enhance collective intelligence*. Nesta Report.

Bouschery, S., Blazevic, V., & Piller, F. (2022). *Artificial intelligence as an actor in hybrid innovation teams: An assessment of the GPT-3 language model*. Forthcoming as a Catalyst Paper in the Journal of Product Innovation Management.

Brauner, P., Dalibor, M., Jarke, M., Kunze, I., Koren, I., Lakemeyer, G., . . . Ziefle, M. (2022). A computer science perspective on digital transformation in production. *ACM Transactions on Internet of Things, 3*(2), 1–32.

Brecher, C., Eckel, H. M., Motschke, T., Fey, M., & Epple, A. (2019). Estimation of the virtual work piece quality by the use of a spindle-integrated process force measurement. *CIRP Annals, 68*(1), 381–384. https://doi.org/hm63

Brecher, C., Özdemir, D., & Weber, A. R. (2016). Integrative production technology: Theory and applications. In C. Brecher & D. Özdemir (Eds.), *Integrative production technology* (pp. 1–17). Springer. https://doi.org/hhn9

Brecher, C., et al. (2017). Learning production systems. In *Proceedings of the 29th AWK Aachener Werkzeugmaschinen-Kolloquium* (pp. 135–161). Apprimus.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., & Askell, A. (2020). *Language models are few-shot learners*. ArXiv:2005.14165.

Cowgill, B., & Tucker, C. E. (2020). *Algorithmic fairness and economics*. Columbia Business School Research Paper.

De Cremer, D. (2020). *Leadership by algorithm: Who leads and who follows in the AI era*. Harriman House.
De Silva, M., Howells, J., & Meyer, M. (2018). Innovation intermediaries and collaboration: Knowledge–based practices and internal value creation. *Research Policy, 47*(1), 70–87. https://doi.org/gcshf5
Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business and Information Systems Engineering, 61*(5), 637–643. https://doi.org/ggkxz4
Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *ArXiv, 1810*, 04805. https://doi.org/hm65
Endsley, M. R. (1987). The application of human factors to the development of expert systems for advanced cockpits. *Proceedings of the Human Factors Society Annual Meeting, 31*(12), 1388–1392. https://doi.org/fzdz4g
Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM, 49*(9), 76–82. https://doi.org/b7f48c
Groensund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *Journal of Strategic Information Systems, 29*(2), 101614. https://doi.org/gjjp64
Hirsch-Kreinsen, H., & Ittermann, P. (2021). Digitalization of work processes: A framework for human-oriented work design. In *The palgrave handbook of workplace innovation* (pp. 273–293). Palgrave Macmillan.
Iansiti, M., & Lakhani, K. R. (2020). Putting AI at the firm's core. *Harvard Business Review, 98*(1), 59–67.
Kamar, E. (2016 July). Directions in hybrid intelligence: Complementing AI systems with human intelligence. In *Proceedings of the twenty-fifth international joint conference on artificial intelligence* (pp. 4070–4073).
Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science, 33*(1), 126–148. https://doi.org/gn3jks
Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 1603–1612).
Liddy, E. D. (2018). Natural language processing for information retrieval. In J. D. McDonald & M. Levine-Clark (Eds.), *Encyclopedia of library and information sciences* (Vol. 5, 4th ed., pp. 3346–3355). CRC Press.
Liebenberg, M., & Jarke, M. (2020). Information systems engineering with digital shadows: Concept and case studies. In S. Dustdar, E. Yu, C. Salinesi, D. Rieu, & V. Pant (Eds.), *Advanced information systems engineering. CAiSE 2020* (Lecture notes in computer science) (Vol. 12127). Springer. https://doi.org/hhph
Long, J. B., & Ehrenfeld, J. M. (2020). The role of augmented intelligence (AI) in detecting and preventing the spread of novel coronavirus. *Journal of Medical Systems, 44*(3), 1–2. https://doi.org/ggp6f3
Mütze-Niewöhner, S., Mayer, C., Harlacher, M., Steireif, N., & Nitsch, V. (2022). Work 4.0: Human-centered work design in the digital age. In W. Frenz (Ed.), *Handbook industry 4.0: Law, technology, society*. Springer.
Pan, Y. (2016). Heading toward artificial intelligence 2.0. *Engineering, 2*(4), 409–413. https://doi.org/gfwwrf
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *Open AI blog, 1*(8), 9.
Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48*(1), 137–141. https://doi.org/ggw7h2
Raj, M., & Seamans, R. (2019). Primer on artificial intelligence and robotics. *Journal of Organization Design, 8*(1), 1–14. https://doi.org/hm67

Shrestha, Y. R., Ben-Menahem, S., & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review, 61*(4), 66–83. https://doi.org/gf48d3

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99–118. https://doi.org/dw3pfg

van der Aalst, W. M. (2016). *Process mining: Data science in action*. Springer.

van der Aalst, W. M. (2020). On the Pareto principle in process mining, task mining, and robotic process automation. In *Proceedings of the 9th international conference on Data Science, Technology and Applications (DATA 2020)* (pp. 5–12). https://doi.org/hm7b

van der Aalst, W. M. (2021). Hybrid Intelligence: To automate or not to automate, that is the question. *International Journal of Information Systems and Project Management, 9*(2), 5–20. https://doi.org/gk92bq

van der Aalst, W. M., Hinz, O., & Weinhardt, C. (2021). Resilient digital twins. *Business and Information Systems Engineering, 63*(6), 615–619. https://doi.org/gmv8sh

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st conference on neural information processing systems*. NIPS.

Von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries, 4*(4), 404–409. https://doi.org/gfztxx

Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In the land of the blind, the one-eyed man is king: Knowledge brokerage in the age of learning algorithms. *Organization Science, 33*(1), 59–82. https://doi.org/gntnhp

Xi, T., Benincá, I. M., Kehne, S., Fey, M., & Brecher, C. (2021). Tool wear monitoring in roughing and finishing processes based on machine internal data. *International Journal of Advanced Manufacturing Technology, 113*(11), 3543–3554. https://doi.org/gndbwx

Zheng, N. N., Liu, Z. Y., Ren, P. J., Ma, Y. Q., Chen, S. T., Yu, S. Y., & Wang, F. Y. (2017). Hybrid-augmented intelligence: Collaboration and cognition. *Frontiers of Information Technology and Electronic Engineering, 18*(2), 153–179. https://doi.org/gg6r35