XAI for Operations in the Process Industry – Applications, Theses, and Research Directions

Arzam Kotriwala[a], Benjamin Kloepper[a], Marcel Dix[a], Gayathri Gopalakrishnan[b], Dawid Ziobro[b] and Andreas Potschka[c]

[a]Industrial Data Analytics, ABB Corporate Research Center, Ladenburg, Germany
[b]User Experience, ABB Corporate Research Center, Vasteras, Sweden
[c]Clausthal University of Technology, Institute of Mathematics, Clausthal-Zellerfeld, Germany

Abstract

Process industry encompasses the transformation of individual raw ingredients into final products. Increasingly, Artificial Intelligence (AI) systems in the industry have led to higher production efficiency, reduced energy consumption, and safer operations. Despite the high degree of automation, human intervention and decision-making remain relevant and important to the required operations. In this contribution, we first present the typical requirements and challenges of applying AI to process industry followed by an overview of Explainable Artificial Intelligence (XAI). Then, we present several theses on successful adoption of XAI for process industry and consequent research gaps and directions. It is shown that the application of XAI in process industry is mainly challenging due to a wide array of requirements arising from a diverse set of AI end-users and AI application cases. An algorithm-centered perspective on XAI research is therefore not enough to address the requirements – future research needs to focus on the interplay between domain knowledge, human factors, and XAI.

Keywords

Explainable AI, Process industry, Manufacturing, User-centered design, Human-automation interaction

1. Introduction

Process industry is the branch of industries that deals with turning input materials (not parts) based on recipes or formulas into products. Examples are oil and gas, chemical, pulp & paper, metal, cement, or food & beverage industries. The operation of these processes deals with the day-to-day matters in the production facility: the monitoring and control of the process, the monitoring and maintenance of equipment, planning and scheduling of the production, and the continuous improvement of the production process, e.g. by recipe changes or improvement of the control processes. The production processes in all these industries are highly automated,

In A. Martin, K. Hinkelmann, H.-G. Fill, A. Gerber, D. Lenat, R. Stolle, F. van Harmelen (Eds.), Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering (AAAI-MAKE 2021) - Stanford University, Palo Alto, California, USA, March 22-24, 2021.

arzam.kotriwala@de.abb.com (A. Kotriwala); benjamin.kloepper@de.abb.com (B. Kloepper); marcel.dix@de.abb.com (M. Dix); gayathri.gopalakrishnan@se.abb.com (G. Gopalakrishnan); dawid.ziobro@se.abb.com (D. Ziobro); andreas.potschka@tu-clausthal.de (A. Potschka)
0000-0002-9099-7689 (A. Kotriwala); 0000-0003-4005-5842 (B. Kloepper); 0000-0001-5984-6594 (M. Dix); 0000-0002-3917-0125 (G. Gopalakrishnan); 0000-0002-7947-772X (D. Ziobro); 0000-0002-6027-616X (A. Potschka)

but human operators, engineers, and maintenance staff still play an essential role [1, 2]. For example, a control room operator ensures productive and effective operations that meet correct product quality while complying with operational and business requirements [3]. Gamer et al. [4] discuss different levels of autonomy in process plants. They point out the contradictory situation that during stable production, plants operate on a high level of autonomy, and in situations like process transition, start-up, shutdowns, or process-upset, there is very little support functionality. However, it is anyhow a popular myth that autonomous systems eliminate the need for Human-Automation interaction. Examples from other domains show that complex deployments of autonomous systems such as military unmanned vehicles, NASA rovers and disaster inspection robots involve people as a critical part of planning and operations [5].

Artificial Intelligence (AI) systems have increasingly proven to be very effective in yielding highly accurate results in several domains. Machine Learning (ML) models that achieve good performance with little false positives and false negative results like Deep Learning networks, Support Vector Machines, or ensemble methods (e.g. Random Forest) are, however, black box models. Besides their primary output (detection of an anomaly, prediction of an event or failure, etc.), they deliver negligible insight on how they achieved their results. Even worse, there are examples of ML models that deliver good performance on training and test sets due to an unknown bias in the available data, but fail to generalize on deployment. This results in two problems: (1) the result of the ML model is not trustworthy, and (2) further investigations to verify, localize, and diagnose the problem that triggered the ML model are required.

Explainable Artificial Intelligence (XAI) is a research area that seeks to address these problems and thus, has increasingly been gaining interest from a wide array of domains. Due to a growing demand for making such opaque systems transparent for better understandability and protection of the rights of end-users [6], XAI has the potential to enable increased adoption and reliability of AI systems. For instance, XAI can help data scientists interpret the inner workings of black-box ML models, data engineers to identify biases in the training data, or justify AI decisions to the domain experts, thereby increasing their trust in the AI solution. In essence, the need for XAI can be broadly categorized into (1) the need for trust and acceptance, and (2) the need for fairness and compliance.

Given the scale and dynamic nature of operations in the process industry, the capacity for teamwork between people and AI systems to ensure reliability and stability of production, is the inevitable next leap forward [7]. As the first step towards teamwork, it is necessary for AI systems to effectively communicate their goals, intentions and conclusions to the people who share the ecosystem. A structured approach towards XAI can help lay the foundations for a future where people work 'with' automation instead of working around automation. Few readers will disagree that deploying ML models for the support of process plant managers and operators promises to yield considerable improvements with respect to safety and process efficiency, and subsequently to reduce the consumption of resources like energy and water.

In this contribution, we highlight typical industrial applications of AI, the data used and the relevant users. Subsequently, we derive research needs and research directions for XAI in the process industry.

Table 1
Examples of AI applications in process industry operations with associated users, data, and methods. (RNN: Recurrent Neural Network; KNN: K-Nearest Neighbor; ANN: Artificial Neural Network; SVM: Support Vector Machine; SVR: Support Vector Regression; RF: Random Forest; IF: Isolation Forest)

Application	End Users	References	AI Methods	Relevant Data
Process monitoring	Operator, Process engineer, Automation engineer	[8, 9, 10]	RNN, KNN	Process signals
Fault diagnosis	Process engineer, Automation engineer, Operator, Maintenance engineer	[11, 12, 13]	ANN, SVM, Bayes Classifier	Process signals, Alarms, Vibration
Event prediction	Operator	[14, 15, 16]	ANN	Process signals, Acoustic signals
Soft sensors	Operator	[17, 18, 19]	SVR, ANN, RF	Process signals
Predictive maintenance	Operator, Maintenance engineer, Scheduler	[20, 21, 22]	RNN, IF	Vibration, Process signals

2. Industrial Applications and Users of AI

Table 1 shows examples of AI applied to use-cases from operational activities in the process industry. This table is not meant as an exhaustive or systematic overview, but should give an indication of the breadth of use cases, users[1], relevant data types, and applied AI methods.

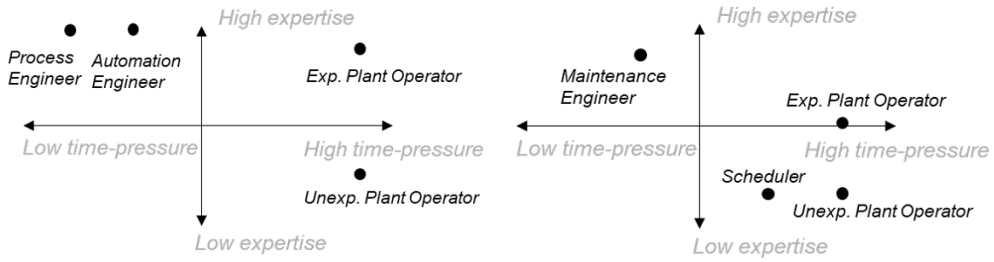
Process monitoring refers to the task of observing the production process in order to detect any problems or upsets. In many plants, time-series data from several hundred signals i.e. process signals, are theoretically available directly from the automation system. Operators observe the process over several different process graphics. Process engineers and automation engineers may even need to monitor even several plants simultaneously. According to the Abnormal Situation Management Consortium Guidelines [23], information overload, inconsistency and inadequacy can all lead to failures in situations involving human supervision. Not surprisingly, constantly observing such high volume of data is a mentally challenging task.

ML solutions can help users assess and react to demanding situations. Once a problem or abnormality in a process or an equipment (e.g. pump or compressor) has been detected, a *fault diagnosis* is performed to identify the cause of the problem. Often, additional sensing such as vibration sensors can enable localization of the problem e.g. leakage in pipe [24]. In fact, some events that require an operator response such as flaring or foaming happen relatively frequently. In order for operators to respond better to or even avoid such events, *event prediction* can help. A *soft sensor* is a term for data-driven methods that provide predicted values for physical or chemical properties that cannot be measured directly or constantly.

Compared to the aforementioned use cases, *predictive maintenance* deals usually with longer time horizons and address a different group of users. The task is to predict the failure of equip-

[1]Not every paper mentions the users explicitly – these are derived from our experience in industrial projects.

Figure 1: Comparison of different users with respect to time-pressure and expertise.
A distinction between experienced (Exp.) and unexperienced (Unexp.) operators is also made.



(a) Fault Diagnosis

(b) Predictive Maintenance

ment in order to plan and schedule maintenance activities with minimal impact on production.

Obviously, the specific requirements of the corresponding users differ across use-cases. Figure 1 illustrates a comparison of the different users in two use-cases with respect to use-case specific expertise of the user and time-pressure demanded by the use-case. For the application *fault diagnosis*, it can be distinguished that the process engineer and automation engineer often deal with problems that affect quality or efficiency on a longer time-horizon and consequently have more time to respond. Plant operators, on the other hand, must typically respond quickly to short-term problems, thereby ensuring safe plant operation. While process and automation engineers have strong theoretic backgrounds, the expertise of the plant operator often heavily depends on individual experience in the specific plant. In the application *predictive maintenance*, the maintenance engineer is interested in long term predictions to plan the maintenance activities and has the expertise to judge the correctness of the machine learning output. Plant operator and scheduler, however, are interested in mid- or short-term failures in order to incorporate this information in their scheduling decisions and control actions respectively.

3. A Brief Overview of XAI

The body of XAI literature is not only vast but also growing at a fast pace with the terms, *explain*, *interpret*, *understand* often used interchangeably [25]. To address the lack of transparency in ML and the consequent need for further verification (see Section 1), the XAI research area, with an overwhelming focus on ML interpretability [26], seeks to make AI systems more human-comprehensible, thereby enabling trust-building as well as compliance.

The task of making ML systems transparent is inherently difficult owing to high complexity of the data and computations involved. In fact, formulating a single comprehensive definition of valid system-human explanations remains challenging [27]. To successfully increase user trust in AI, the justification provided must match the domain as well as the complexity or understandability of the user [6]. This has led researchers to address this challenge from different perspectives, resulting in diverse XAI approaches and consequent explanation types. Several attempts to provide a taxonomy for XAI methods have already been made (e.g. [28, 29]). Based on these classifications, the most distinct properties of XAI methods are summarized in Table 2.

Table 2
Summary of important properties of XAI approaches from literature.

Property	Types
Interpretability	Intrinsic, post-hoc
Scope	Global, local
Mode	Interactive, static
Supported data types	Tabular, image, text, time-series
Mechanism	Attribution, surrogate, similarity, prototype, counterfactual

Since there is no one XAI approach that works for all users [28], depending on specific use-case requirements, these properties can guide selection of the most appropriate approaches.

The majority of XAI research has taken an algorithmic focus [30]. While some ML methods, such as linear regression, are intrinsically transparent, most XAI methods provide interpretability post-hoc i.e. as an auxiliary component [31]. The methods may also differ in scope – the model may be explained (global) or a specific prediction (local) or both. The state-of-the-art explanation mechanisms can be broadly categorized into feature attribution methods, surrogate models, counterfactuals, representative examples of classes (prototypes), case-based explanations, causal mechanism (e.g. rules), textual or simply visual. These representations are, however, not free from overlap – prototypes, counter examples, or feature attributions, for instance, may be presented in a visual way. Most of these XAI techniques overwhelmingly cater to specific data types such as images, and thus when applied to time-series data, do not fully facilitate increased human understanding [32]. For instance, despite LIME [33] being a popular model-agnostic XAI technique, it was shown to yield poor performance on time-series data, most likely owing to high dimensionality of the data and its use of a linear classifier. The shortcomings of these XAI methods for temporal data make them inherently difficult to apply to most use-cases arising in the process industry.

4. Theses on XAI for Process Industry

This section puts forward theses about the relevance and successful adoption of XAI in the process industry and the respective requirements from the application domain. In this contribution, they are presented as theses as they require further empirical validation.

Thesis 1: Explainability is critical
Safety is paramount in the process industry and hence, the industry is very risk-aware [34]. The black-box character of many of the methods presented in Table 1, such as deep neural networks or ensemble methods, is one of the major obstacles in the application of AI technologies. It is a known problem in the industry that operators may lose confidence in model-predictive control and deactivate such solutions [35]. Black-box machine learning models without explanations will very likely share a similar fate. Therefore, for successful adoption of AI systems, valid explanations must be offered that satisfy the industry expert's need to justify and prove resulting decisions taken [7].

Thesis 2: Local explanations are highly relevant
XAI for operations in the process industry should help the end-user of ML to understand the model and draw conclusions from the prediction. With respect to the available time for decision-making and use-case expertise in most applications (see Figure 1), the need for local explanations of predictions is more pronounced. In most cases, ML will not close the loop directly, i.e., ML will not take decisions and automatically trigger their implementation. Instead, ML mostly takes the role of a decision support system in industrial use-cases, either highlighting relevant information or recommending a specific decision to the human. The responsibility to take the decision stays with the human.

Thesis 3: The choice of mechanism is key
The explanations of individual predictions should enable the human to validate the correctness of the output (e.g., whether there is going to be a quality problem or not) as well as to draw the right conclusion (e.g. selecting an appropriate action). In many use-cases and for many users, these two processes need to be performed under considerable time-pressure. When deciding on a XAI technique, two factors are important: the time to create an explanation and the time it takes the user to understand the explanation. Whilst the first metric is discussed in some XAI research, the authors know only one XAI publication [36] that evaluates the time to understand the explanation. Humans have limited resources in perceptual modalities and currently, there is a gap in understanding how explainability techniques should complement the system-human communication instead of interfering with it. For example, there has been research on perceptual modalities indicating that people sometimes divide attention between the eye and ear better than between two auditory or two visual ones [37]. Understanding which types of explanations users can comprehend under time-pressure and the appropriate modalities for the explanation are important aspects when designing XAI systems for the process industries.

Thesis 4: Dynamic and tailored explanations are needed
Suitable explanations are not only dependent on the time available to AI users for decision-making but also on the AI prediction output and the specific application situation. A static explanation, that always presents the same explanation regardless of the user's context, cannot meet these requirements. One way to address this might be via interactive explanations that allow users to move from simple, high-level explanations to more detailed explanations, featuring different types of explanations, will be very beneficial in the industrial context. However, surprisingly, increased transparency may even hamper people's ability to detect when the model makes a sizable mistake and correct for it, seemingly due to information overload [38]. An alternative to dynamic explanations might be to provide non-interactive explanation, but to choose the type and level of detail of the explanation depending on the user's context.

Thesis 5: Domain expert and knowledge must take center-stage
In the process industry, for safety and reliability of production, domain expert users have the responsibility to ensure compliance to industry standards [7]. Thus, the role of AI is largely to support industrial users in making their final decisions using their expertise and situational knowledge. Consequently, it is important for AI solutions to not only provide explanations but

also to equip the domain experts with appropriate tools and interfaces to provide feedback to and modify the AI system, thereby retaining their control. In fact, many view an explanation as a dialog that enables the user and system to achieve a shared understanding [39]. Such a shared understanding may also help codify expert knowledge thereby potentially fostering standardization in the plant and knowledge transfer to the new workforce.

5. Suggested Research Directions

Based on the aforementioned theses, we suggest three directions of research for XAI in the process industry: user-centered research as an important activity to validate and refine several of the theses brought forward in the previous section; research on specific algorithms and methods that address the needs discussed within the theses; and explanations that are dynamic and derived using multi-mechanisms, in order to cater the varying needs of different users.

5.1. User-Centered Approach to XAI

In the design of automated systems in the process industry, traditionally, the automation is at the heart of the system with the expectation that the users will adapt to the automation. It is well known that the success of this approach is limited [40, 41]. Advanced automation does not necessarily improve operator performance [42]. Human-Centered AI is a compelling prospect that enables people to see, think, create, and act in extraordinary ways, by combining potent user experiences with embedded AI methods to support services that users want [43]. It reverses the narrative and treats the users' needs, goals, and capabilities as the core around which automation is built.

The act of explanation is inherently social [44]. Irrespective of the explainer being human or an algorithm (as in XAI), the explanation must be adapted to the context of the recipients for a successful communication of the explanation. If an explanation is not meaningful to the user, their perception of the system might be affected negatively [45]. Explanations with too few details and too much detail can cause users to lose trust in the system [46]. Even the need for explanations is context dependent. In some cases, explanations have no impact on decision making [47]. To understand what qualifies as a meaningful explanation to the user, we need to understand the user and their context.

Methods such as mental model elicitation [48], Cognitive Task Analysis and contextual inquiry help us understand how expert users assimilate information and make decisions [49]. Co-creation and participatory design approaches can help customize explanation to specific domain. A user-centered approach rooted in human factors, cognitive science and user experience can help engineer user-friendly AI solutions for process industries.

5.2. XAI Methods for Industrial Data

Several popular XAI methods such as SHAP Values [50], LIME [33], or feature importance plots (e.g., [51]) provide explanations by feature attribution - identifying the features with the highest relevance. The application of these methods to multi-variate signal data, possibly of variable length, which is common in process industry applications, is challenging. Although

there exist applications of SHAP and LIME to time-series data, they are typically suited to univariate time-series [52]. Not surprisingly, an evaluation of XAI methods for time-series data [32] also raises the need for more abstract representations and to develop more sophisticated approaches to XAI for time-series data.

Both SHAP and feature importance plots rely on varying the input features across the feature distribution obtained from the data set. Naively applying this to signal data - and individually changing every point in each signal, will yield samples that are unrealistic or even infeasible. How that will impact the reliability of the feature attribution remains unclear. LIME and the related method, Anchors [36] use feature perturbation, varying features in accordance to the mechanisms of the data-producing domain. However, the question of how to obtain good feature perturbation distributions for industrial processes that are also within the time requirements is an open research question.

Explainability based on case-based-reasoning or prototypes [53, 54] appears to be a better approach to support ML usage in the process industry. However, the authors are not aware of any application to multi-variate time-series or even process industry. Model-specific methods like saliency maps for deep learning networks [55] or shapelet-based explanations [56, 57] for random forest are interesting methods, if the corresponding ML models are used. For use-cases in process industry, shapelet-based techniques could be a very interesting approach to create global surrogate models (for instance, based on decision trees).

Developing approaches that embed domain expertise into ML pipelines (such as TED [6]) can be very beneficial in the process industry. Whilst such techniques may be difficult to directly scale to larger datasets, they have the advantage that the resulting explanations are engineered by the domain experts themselves and are thus, likely to be more meaningful to them.

5.3. Dynamic and Multi-Mechanism Explanations

Different users have varied requirements that are influenced by factors such as time-pressure and experience. AI solutions need to be put into context for different users and a one-fits-all AI solution is therefore not sufficient. Dynamic or interactive explanations should allow users to perform drill-downs or to choose from different explanation mechanisms. According to [58], this type of XAI is, in general, a white spot in the research landscape. A few examples are [59] and [31], that mainly discuss requirements. The development of dynamic explanations that also support or contrast predictions with examples from the historical data is an important research direction with relevance beyond the application area of the process industry.

6. Conclusion

This contribution has introduced the domain of process industry operations as a challenging research field for the application of XAI. The field is signified by diverse AI application cases and end-users who, with varied requirements and expertise, play an essential role in safe and reliable process operations. These industrial experts require justifications which match their domain knowledge to support them in making critical decisions and remaining compliant to industrial standards. When the models are explainable, the AI end-users will be assured that outcomes are bias-free, safe, legal, ethical and appropriate for industrial settings.

We propose that a multi-disciplinary approach should be taken for successful and sustainable application of XAI to operations in the process industry. Special consideration needs to be given to understanding how domain experts operate and to include them in validation of XAI methods, which also need to be specifically tailored for industrial data. This will require the cooperation of researchers from AI, user experience, psychology, and process industry experts.

Acknowledgement

The authors thank Divya Sheel (ABB Corporate Research Center, India) for reviewing the paper.

References

[1] G. Bello, V. Colombari, The human factors in risk analyses of process plants: The control room operator model 'teseo', Reliability engineering 1 (1980) 3–14.

[2] L. H. Ikuma, C. Harvey, C. F. Taylor, C. Handal, A guide for assessing control room operator performance using speed and accuracy, perceived workload, situation awareness, and eye tracking, Journal of loss prevention in the process industries 32 (2014) 454–465.

[3] S. Nazir, S. Colombo, D. Manca, The role of situation awareness for the operators of process industry, Chemical Engineering Transactions 26 (2012).

[4] T. Gamer, M. Hoernicke, B. Kloepper, R. Bauer, A. J. Isaksson, The autonomous industrial plant-future of process engineering, operations and maintenance, IFAC-PapersOnLine 52 (2019) 454–460.

[5] J. M. Bradshaw, R. R. Hoffman, D. D. Woods, M. Johnson, The seven deadly myths of" autonomous systems", IEEE Intelligent Systems 28 (2013) 54–61.

[6] M. Hind, D. Wei, M. Campbell, N. C. Codella, A. Dhurandhar, A. Mojsilović, K. Natesan Ramamurthy, K. R. Varshney, Ted: Teaching ai to explain its decisions, in: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 2019, pp. 123–129.

[7] M. W. Hoffmann, R. Drath, C. Ganz, Proposal for requirements on industrial ai solutions, in: Machine Learning for Cyber Physical Systems, Springer Vieweg, Berlin, Heidelberg, 2020, pp. 63–72.

[8] B. Mamandipoor, M. Majd, S. Sheikhalishahi, C. Modena, V. Osmani, Monitoring and detecting faults in wastewater treatment plants using deep learning, Environmental Monitoring and Assessment 192 (2020) 148.

[9] I. M. Cecílio, J. R. Ottewill, J. Pretlove, N. F. Thornhill, Nearest neighbors method for detecting transient disturbances in process and electromechanical systems, Journal of Process Control 24 (2014) 1382–1393.

[10] L. Banjanovic-Mehmedovic, A. Hajdarevic, M. Kantardzic, F. Mehmedovic, I. Dzananovic, Neural network-based data-driven modelling of anomaly detection in thermal power plant, Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije 58 (2017) 69–79.

[11] I. Yélamos, M. Graells, L. Puigjaner, G. Escudero, Simultaneous fault diagnosis in chemical plants using a multilabel approach, AIChE Journal 53 (2007) 2871–2884.

[12] M. Lucke, A. Stief, M. Chioua, J. R. Ottewill, N. F. Thornhill, Fault detection and identification combining process measurements and statistical alarms, Control Engineering Practice 94 (2020) 104195.
[13] D. Ruiz, J. Canton, J. M. Nougués, A. Espuna, L. Puigjaner, On-line fault diagnosis system support for reactive scheduling in multipurpose batch chemical plants, Computers & Chemical Engineering 25 (2001) 829–837.
[14] G. Dorgo, P. Pigler, M. Haragovics, J. Abonyi, Learning operation strategies from alarm management systems by temporal pattern mining and deep learning, in: Computer Aided Chemical Engineering, volume 43, Elsevier, 2018, pp. 1003–1008.
[15] M. Giuliani, G. Camarda, M. Montini, L. Cadei, A. Bianco, A. Shokry, P. Baraldi, E. Zio, et al., Flaring events prediction and prevention through advanced big data analytics and machine learning algorithms, in: Offshore Mediterranean Conference and Exhibition, Offshore Mediterranean Conference, 2019.
[16] A. Carter, L. Briens, An application of deep learning to detect process upset during pharmaceutical manufacturing using passive acoustic emissions, International journal of pharmaceutics 552 (2018) 235–240.
[17] K. Desai, Y. Badhe, S. S. Tambe, B. D. Kulkarni, Soft-sensor development for fed-batch bioreactors using support vector regression, Biochemical Engineering Journal 27 (2006) 225–239.
[18] C. Shang, F. Yang, D. Huang, W. Lyu, Data-driven soft sensor development based on deep learning technique, Journal of Process Control 24 (2014) 223–233.
[19] L. F. Napier, C. Aldrich, An isamill™ soft sensor based on random forests and principal component analysis, IFAC-PapersOnLine 50 (2017) 1175–1180.
[20] I. Amihai, R. Gitzel, A. M. Kotriwala, D. Pareschi, S. Subbiah, G. Sosale, An industrial case study using vibration data and machine learning to predict asset health, in: 2018 IEEE 20th Conference on Business Informatics (CBI), volume 1, IEEE, 2018, pp. 178–185.
[21] I. Amihai, M. Chioua, R. Gitzel, A. M. Kotriwala, D. Pareschi, G. Sosale, S. Subbiah, Modeling machine health using gated recurrent units with entity embeddings and k-means clustering, in: 2018 IEEE 16th International Conference on Industrial Informatics (INDIN), IEEE, 2018, pp. 212–217.
[22] N. Kolokas, T. Vafeiadis, D. Ioannidis, D. Tzovaras, Fault prognostics in industrial domains using unsupervised machine learning classifiers, Simulation Modelling Practice and Theory (2020) 102109.
[23] P. Bullemer, Effective console operator HMI design, ASM Consortium, Houston, TX, 2015.
[24] M. Hollender, Ai-supported workflows for chemical batch plants: Optimizing quality, efficiency and safety, atp magazin 62 (2020) 84–88.
[25] D. Doran, S. Schulz, T. R. Besold, What does explainable ai really mean? a new conceptualization of perspectives, arXiv preprint arXiv:1710.00794 (2017).
[26] D. V. Carvalho, E. M. Pereira, J. S. Cardoso, Machine learning interpretability: A survey on methods and metrics, Electronics 8 (2019) 832.
[27] Z. C. Lipton, The mythos of model interpretability, Queue 16 (2018) 31–57.
[28] V. Arya, R. K. Bellamy, P.-Y. Chen, A. Dhurandhar, M. Hind, S. C. Hoffman, S. Houde, Q. V. Liao, R. Luss, A. Mojsilović, et al., One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques, arXiv preprint arXiv:1909.03012 (2019).

[29] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, arXiv (2019) arXiv–1910.
[30] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, Artificial Intelligence 267 (2019) 1–38.
[31] K. Sokol, P. Flach, One explanation does not fit all, Künstliche Intelligenz 34 (2020) 235–250.
[32] U. Schlegel, H. Arnout, M. El-Assady, D. Oelke, D. A. Keim, Towards a rigorous evaluation of xai methods on time series, arXiv preprint arXiv:1909.07082 (2019).
[33] M. T. Ribeiro, S. Singh, C. Guestrin, " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.
[34] E. De Rademaeker, G. Suter, H. J. Pasman, B. Fabiano, A review of the past, present and future of the european loss prevention and safety promotion in the process industries, Process Safety and Environmental Protection 92 (2014) 280–291.
[35] M. G. Forbes, R. S. Patwardhan, H. Hamadah, R. B. Gopaluni, Model predictive control in industry: Challenges and opportunities, IFAC-PapersOnLine 48 (2015) 531–538.
[36] M. T. Ribeiro, S. Singh, C. Guestrin, Anchors: High-precision model-agnostic explanations., in: AAAI, volume 18, 2018, pp. 1527–1535.
[37] C. D. Wickens, J. G. Hollands, S. Banbury, R. Parasuraman, Engineering psychology and human performance, Psychology Press, 2015.
[38] F. Poursabzi-Sangdeh, D. G. Goldstein, J. M. Hofman, J. W. Vaughan, H. Wallach, Manipulating and measuring model interpretability, arXiv preprint arXiv:1802.07810 (2018).
[39] S. T. Mueller, R. R. Hoffman, W. Clancey, A. Emrey, G. Klein, Explanation in human-ai systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable ai, arXiv preprint arXiv:1902.01876 (2019).
[40] L. Bainbridge, Ironies of automation, in: Analysis, design and evaluation of man–machine systems, Elsevier, 1983, pp. 129–135.
[41] R. I. Cook, How complex systems fail, Cognitive Technologies Laboratory, University of Chicago. Chicago IL (1998).
[42] C. D. Wickens, H. Li, A. Santamaria, A. Sebok, N. B. Sarter, Stages and levels of automation: An integrated meta-analysis, in: Proceedings of the human factors and ergonomics society annual meeting, volume 54, Sage Publications Sage CA: Los Angeles, CA, 2010, pp. 389–393.
[43] B. Shneiderman, Human-centered artificial intelligence: Three fresh ideas, AIS Transactions on Human-Computer Interaction 12 (2020) 109–124.
[44] O. Biran, C. Cotton, Explanation and justification in machine learning: A survey, in: IJCAI-17 workshop on explainable AI (XAI), volume 8, 2017, pp. 8–13.
[45] M. Nourani, S. Kabir, S. Mohseni, E. D. Ragan, The effects of meaningful and meaningless explanations on trust and perceived system accuracy in intelligent systems, in: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, volume 7, 2019, pp. 97–105.
[46] R. F. Kizilcec, How much information? effects of transparency on trust in an algorithmic

interface, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016, pp. 2390–2395.

[47] Y. Alufaisan, L. R. Marusich, J. Z. Bakdash, Y. Zhou, M. Kantarcioglu, Does explainable artificial intelligence improve human decision-making?, arXiv preprint arXiv:2006.11194 (2020).

[48] J. Dodge, S. Penney, C. Hilderbrand, A. Anderson, M. Burnett, How the experts do it: Assessing and explaining agent behaviors in real-time strategy games, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1–12.

[49] H. Mucha, S. Robert, R. Breitschwerdt, M. Fellmann, Towards participatory design spaces for explainable ai interfaces in expert domains, 43rd German Conference on Artificial Intelligence, Bamberg, Germany (2020).

[50] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[51] A. Goldstein, A. Kapelner, J. Bleich, E. Pitkin, Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation, Journal of Computational and Graphical Statistics 24 (2015) 44–65.

[52] E. Metzenthin, Lime-for-time, 2017. URL: https://github.com/emanuel-metzenthin/Lime-For-Time.

[53] O. Li, H. Liu, C. Chen, C. Rudin, Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions, arXiv preprint arXiv:1710.04806 (2017).

[54] E. M. Kenny, M. T. Keane, Twin-systems to explain artificial neural networks using case-based reasoning: comparative tests of feature-weighting methods in ann-cbr twins for xai, in: Twenty-Eighth International Joint Conferences on Artifical Intelligence (IJCAI), Macao, 10-16 August 2019, 2019, pp. 2708–2715.

[55] T. N. Mundhenk, B. Y. Chen, G. Friedland, Efficient saliency maps for explainable ai, arXiv preprint arXiv:1911.11293 (2019).

[56] L. Ye, E. Keogh, Time series shapelets: a new primitive for data mining, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009, pp. 947–956.

[57] I. Karlsson, J. Rebane, P. Papapetrou, A. Gionis, Explainable time series tweaking via irreversible and reversible temporal transformations, in: 2018 IEEE International Conference on Data Mining (ICDM), IEEE, 2018, pp. 207–216.

[58] V. Arya, R. K. E. Bellamy, P.-Y. Chen, A. Dhurandhar, M. Hind, S. C. Hoffman, S. Houde, Q. V. Liao, R. Luss, A. MojsiloviÄ‡, S. Mourad, P. Pedemonte, R. Raghavendra, J. T. Richards, P. Sattigeri, K. Shanmugam, M. Singh, K. R. Varshney, D. Wei, Y. Zhang, Ai explainability 360: An extensible toolkit for understanding data and machine learning models, Journal of Machine Learning Research 21 (2020) 1–6.

[59] M. Chromik, reshape: A framework for interactive explanations in xai based on shap, in: Proceedings of 18th European Conference on Computer-Supported Cooperative Work, European Society for Socially Embedded Technologies (EUSSET), 2020.