Machine Learning in Chemical Engineering

A Perspective

Schweidtmann, Artur M.; Esche, Erik; Fischer, Asja; Kloft, Marius; Repke, Jens Uwe; Sager, Sebastian; Mitsos, Alexander

DOI
10.1002/cite.202100083

Publication date
2021

Document Version
Final published version

Published in
Chemie-Ingenieur-Technik

Citation (APA)
Schweidtmann, A. M., Esche, E., Fischer, A., Kloft, M., Repke, J. U., Sager, S., & Mitsos, A. (2021). Machine Learning in Chemical Engineering: A Perspective. *Chemie-Ingenieur-Technik*, *93*(12), 2029-2039. https://doi.org/10.1002/cite.202100083

Machine Learning in Chemical Engineering: A Perspective

Artur M. Schweidtmann[1,2,*], Erik Esche[3], Asja Fischer[4], Marius Kloft[5], Jens-Uwe Repke[3], Sebastian Sager[6], and Alexander Mitsos[2,7,8]

DOI: 10.1002/cite.202100083

The transformation of the chemical industry to renewable energy and feedstock supply requires new paradigms for the design of flexible plants, (bio-)catalysts, and functional materials. Recent breakthroughs in machine learning (ML) provide unique opportunities, but only joint interdisciplinary research between the ML and chemical engineering (CE) communities will unfold the full potential. We identify six challenges that will open new methods for CE and formulate new types of problems for ML: (1) optimal decision making, (2) introducing and enforcing physics in ML, (3) information and knowledge representation, (4) heterogeneity of data, (5) safety and trust in ML applications, and (6) creativity. Under the umbrella of these challenges, we discuss perspectives for future interdisciplinary research that will enable the transformation of CE.

Keywords: Deep learning, Hybrid modeling, Machine learning, Optimization, Reinforcement learning

Received: May 28, 2021; *revised:* August 25, 2021; *accepted:* October 12, 2021

1 Introduction

The chemical industry must convert to using renewable energy and feedstock supply, otherwise chemical production might become the largest driver of global oil consumption by 2030 [1–4]. However, renewable resources fluctuate over time and space, requiring dynamic operation and a new paradigm for identifying new process routes and the design of flexible plants [3]. At the same time, the chemical companies are facing increased competition and must ensure optimal operation and short development cycles for new processes. Facilitating this radical change poses difficulties as conventional methods for process synthesis and operation may not be sufficient. To make optimal decisions in complex environments, models are conventionally developed based on mechanistic understanding and optimized. However, the development of physicochemical models is expensive, and many phenomena cannot be fully described by computationally tractable models.

Machine learning (ML) has the potential to overcome the limitations of mechanistic modeling as ML methods can learn complex behaviors, the model development is cheaper, and it can be advantageous for optimization [5, 6]. Chemical engineering (CE) already experienced two big waves of ML applications between the 1980s and 2008, i.e., expert systems and (shallow) artificial neural networks (ANNs) (c.f. [7, 8]). These waves had limited impact due to several reasons [7]: i) lack of data, of data accessibility, of computation power, and of programming environments/paradigms, ii) competing successful emerging technologies for CE, in particular, mechanistic modeling, optimization, and model predictive control.

[1]Prof. Artur M. Schweidtmann
a.schweidtmann@tudelft.nl
Delft University of Technology, Department of Chemical Engineering, Van der Maasweg 9, 2629 HZ Delft, The Netherlands.

[2]Prof. Artur M. Schweidtmann, Prof. Alexander Mitsos, Ph.D.
RWTH Aachen University, Aachener Verfahrenstechnik, Forckenbeckstr. 51, 52074 Aachen, Germany.

[3]Dr. Erik Esche, Prof. Dr. Jens-Uwe Repke
Technische Universität Berlin, Fachgebiet Dynamik und Betrieb technischer Anlagen, Straße des 17. Juni 135, 10623 Berlin, Germany.

[4]Prof. Dr. Asja Fischer
Ruhr-Universität Bochum, Department of Mathematics, Universitätsstraße 150, 44801 Bochum, Germany.

[5]Prof. Dr. Marius Kloft
Technische Universität Kaiserslautern, Department of Computer Science, Erwin-Schrödinger-Straße 52, 67663 Kaiserslautern, Germany.

[6]Prof. Dr. Sebastian Sager
Otto-von-Guericke-Universität Magdeburg, Department of Mathematics, Universitätsplatz 2, 39106 Magdeburg, Germany.

[7]Prof. Alexander Mitsos, Ph.D.
JARA Center for Simulation and Data Science (CSD), Aachen, Germany.

[8]Prof. Alexander Mitsos, Ph.D.
Forschungszentrum Jülich, Institute for Energy and Climate Research IEK-10 Energy Systems Engineering, Wilhelm-Johnen-Straße, 52428 Jülich, Germany.

Today, we have cheap and powerful computing, easy-to-use programming environments (e.g., Python & TensorFlow), and a large open-source community in ML. At the same time, ML has seen a surge in automatic feature learning by deep ANNs [9, 10]. This development together with advances in hardware – most importantly GPU computing – led to breakthrough results in image recognition [11], especially when based on convolutional neural networks (CNNs) [7], and in game playing [12–14] and generally to a technology push in ML [15].

CE currently undergoes a transformation towards digitization and full automation of industry and research. This leads to an ever-increasing availability of data and the need for automated optimal decision-making based on data, allowing for more sustainable process operations [16]. We thus have a technology push and industry pull situation, where ML opens up new possibilities to overcome pressing challenges in CE [7]. In this perspective, we first review ML methods already established in CE (Sect. 2). Then, we identify six emerging ML challenges with great potentials in CE (Sect. 3).

2 Established Machine Learning Methods in Chemical Engineering

ML is a subclass of artificial intelligence (AI). ML has roots in computer science and mathematics and gives computers the ability to learn from data without being explicitly programmed. ML is broadly classified into supervised learning and unsupervised learning [17]. Other types of ML are reinforcement learning (RL) as well as hybrids such as semi-supervised learning.

First applications of classical AI in CE were proposed in the 1980s with the advent of expert systems, e.g., for thermophysical properties [18] and catalyst design [19]. They did not achieve breakthroughs mainly because implementation, training, and maintenance were costly and time-consuming [7] (c.f. Sect. 1). As ML theory, computer hardware, and programming languages advanced, ML was applied to experimental and simulated data to extract information, recognize patterns, and make predictions [20]. Overall, ML methods for process monitoring, fault detection, and soft sensing are mostly mature and commercially available in CE.

2.1 Unsupervised Learning

Unsupervised learning describes the collection of techniques that investigate "unlabeled data", i.e., data with no explicit input-output connection. The main purpose is to find hidden structure in data, e.g., for clustering, feature extraction, compression, or anomaly detection. Unsupervised learning is popular and attractive from a practical point of view as input-output connections are oftentimes unavailable in applications.

In CE, process monitoring and fault detection have seen many applications over the past decades leading to commercial tools and industrial applications. Process monitoring is mainly based on classical principal component analysis (PCA) [21], while some researchers investigated independent component analysis for non-Gaussian processes [22] and kernel density estimation for applications with unknown distributions, e.g., for data smoothing [23]. Further advances are monitoring platforms, e.g., using self-organizing maps for a wastewater treatment plant [24] and Gaussian mixture models for the Tennessee Eastman process [25]. Fault detection is another common application of unsupervised learning to process data. In the previous literature, variations of PCA have been used frequently for fault detection [26–28]. Furthermore, other advanced methods have been applied to distinguish between normal and faulty batches (e.g., support vector data description [29] and k-means clustering [30]). Today, first fault detection tools are commercially available for the process industry.

2.2 Supervised Learning

Supervised learning methods train a model on labeled data with an explicit input-output structure and learn functions mapping an input to an output. Regression is a supervised ML tool that is part of the standard repertoire in process systems engineering (PSE) and has long been used for modeling and subsequent optimal design of processes. Regarding soft sensor applications, i.e., online prediction of process qualities, a large variety of different methods has been used, including partial least squares [31–33], principal component regression [34, 35], support vector machines (SVM) [36], ANNs [37], and Gaussian process (GP) regression [38–40]. Applications of these include complex large-scale processes such as air separation units [32], injection-molding [35], reverse osmosis of seawater [33], and further chemical production processes [37, 38, 41].

Supervised learning has long been used for dynamic systems in operations and control. A wide range of models is applied to describe dynamic processes based on data in discrete-time and continuous-time approaches [42]. There are state-space models, Hammerstein-Wiener models, scale-bridging surrogate models, linear autoregressive integrated moving average (ARIMAX), and nonlinear ARIMAX (NARMAX) [43]. Identification of these models is well established. However, they are limited to Markovian systems, where the current state completely describes the system, i.e., effects of hysteresis cannot be described. In ML, recurrent neural networks (RNN) have been introduced to include non-Markovian effects. For example, [44] use an RNN to learn the policy for operating a batch bioprocess. In cases where large data sets are present and long-term dependencies are relevant, training of standard RNNs suffers from vanishing gradients and gated recurrent neural

networks like long-short-term memory (LSTM) architectures are suitable [45].

3 Emerging Machine Learning Challenges in Chemical Engineering

Beyond the previously summarized topics deeply rooted in data mining and analytics, we identify six emerging challenges of ML with a large potential for CE [46].

3.1 Optimal Decision Making

Optimal decision making is a prominent topic in CE, for process synthesis, control, as well as solvent, catalyst, or adsorbent selection. All these decisions need to be made based on existing information which can be in the form of data and mechanistic knowledge, e.g., models. As shown in Fig. 1, optimal decision making based on data can be done by training of data-driven or hybrid models and subsequent optimization with embedded data-driven models [5, 6].

Within PSE, various data-driven models have been used for regression and subsequent process optimization. As illustrated in Fig. 2, the complexity of applied data-driven models ranges from linear approximations to deep ANNs. For a long time, the literature focused on linear models to approximate simulation and experimental data [47]. Since the 1990s, shallow ANNs have been used extensively. Shallow ANNs can theoretically approximate any nonlinear, smooth function given a sufficient number of neurons to any given positive accuracy on a training data set [48]. In many PSE applications, ANNs are fitted to complete processes (black-box approach, e.g., [49]), or ANNs are combined with mechanistic model equations (hybrid modeling approach, e.g., [50–52]). Subsequently, the obtained process models can be optimized, e.g., to identify process design [53]. Today, the re-emergence of ML is mostly driven by deep ANNs and big data [15]. The deep ANNs are believed to become more important in PSE because abundant data becomes available, e.g., through smart manufacturing, high throughput experiments, and simulation studies [16]. However, applications of deep ANNs are still limited in process design in PSE [8].

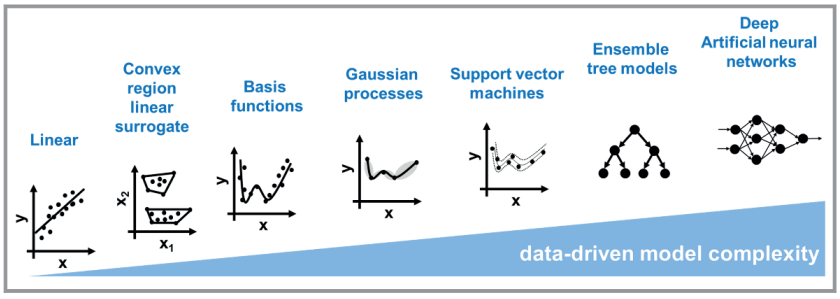


Figure 2. Overview of data-driven models embedded in optimization problems in CE: linear [47], convex region linear surrogate model [57], nonlinear basis functions, e.g., ALAMO [56], piecewise polynomial function [59], spline function [60], Gaussian process [54, 55], support vector machine, ensemble tree model (e.g., random forest, gradient boosted trees [58, 64]), ANNs with ReLU activation [61–63], and other ANNs with more complex activation functions [6, 54]. Note that the models are ordered by their estimated ability to learn complex dependencies.

While optimization of problems with linear models can be solved globally on a large scale, e.g., for structural optimization [47], linear models cannot learn high-dimensional nonlinear problems accurately. On the other hand, the consideration of more complex data-driven models like ANNs and GPs, which could reflect high dimensional nonlinear problems better, has long been limited to local or stochastic solution approaches [51–55]. Although some tailor-made data-driven models can be solved using state-of-the-art global solvers, these are also limited to low-dimensional problems [56, 57]. A few researchers in CE and ML have developed tailored optimization approaches for problems with ML models embedded. Mistry et al. [58] proposed a tailored algorithm for problems with gradient boosted trees embedded. Grimstad and coworkers proposed an algorithm for the optimization of piecewise polynomial functions [59] and spline functions [60]. Some previous works also used general-purpose global solvers to solve optimization problems with complex surrogate models embedded [54, 55] but observed computational burdens. Recently, Schweidtmann and Mitsos [6] proposed an efficient reduced-space optimization formulation for global optimization of deep ANNs. Notably, ANNs with rectified linear unit (ReLU) activations have recently been reformulated as mixed-integer linear programs (MILPs) [61–63]. In the MILP formulations, binary variables are introduced to divide the domain of the piecewise linear ReLU activation functions into two linear sub-domains. Similarly, tree models can be reformulated as MILPs [58, 64, 65]. However, the number of integer variables and constraints grows linearly with the model complexity (e.g., number of nodes in the ANN).

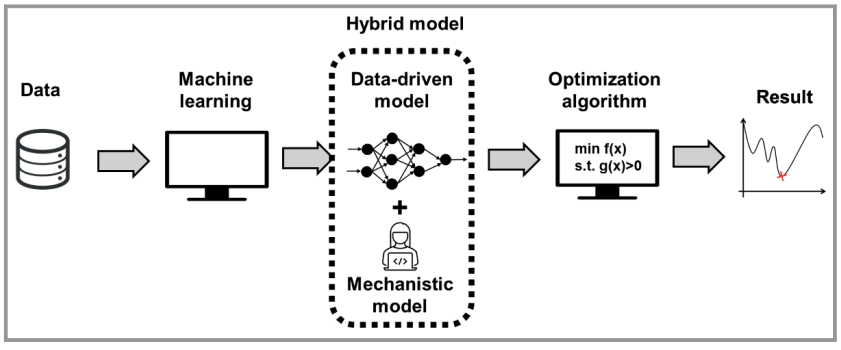


Figure 1. Illustration of the data-driven modeling and optimization approach [5, 6].

Overall, optimal decision-making based on data has already seen some work on optimal process synthesis and optimal process operation. However, there are still severe limitations in integrating learning and optimization frameworks that exhibit complex ML models.

Promising future research in optimal decision making includes ML-assisted/embedded optimization and ML-assisted control. In both areas, the complexity of real-life processes and the inclusion of non-Markovian effects promises to discover insufficiencies of ML methods, to nurture new developments, and to open promising new avenues of research in both fields. This area has excellent synergy potential with the other five.

3.2 Introducing and Enforcing Physics in Machine Learning

The physicality of ML models is a frequently called-for development in CE. Across many disciplines, supervised learning is directly applied in a black-box approach. However, black-box approaches have severe drawbacks in interpretability, extrapolation, data demand, and reliability. These drawbacks can limit applications of ML and lead to fatal errors when being applied in industry without necessary checks. On the other hand, mechanistic, physicochemical models can provide structural knowledge that can be combined with data-driven models.

The combination of mechanistic and data-driven models is called hybrid (semi-parametric) modeling. It promises advantages such as better interpretability, enhanced extrapolation properties, and higher prediction accuracy [66–68]. Fig. 3 illustrates the enhanced extrapolation properties of hybrid models. In the illustrative example, some training data points are distributed on low-dimensional manifolds that are illustrated by dashed lines. In this case, a standard black-box approach (e.g., an ANN with two inputs and one output) should not be evaluated outside the manifolds of the training data points. In the illustrative case, a mechanistic model structure $g(f_1(x_1)f_2(x_2))$ is known a priori where f_1 and f_2 are unknown functions and g is a known mechanistic model. This allows to build a hybrid model that can be evaluated outside the initial training data manifold because each black-box model has only one single input and is thus evaluated within its training data range. Consequently, the hybrid model structure avoids extrapolation of the black-box model parts. Hybrid modeling has numerous applications in CE and biotechnology since the early 1990s, e.g., in process [69, 70] and reactor modeling [71], polymerization [72], crystallization, distillation, drying processes [73], and process control [74, 75]. Also, many empirical constitutive equations in CE can be interpreted as simple data-driven parts in a hybrid model. Hybrid modeling has strong theoretical foundations within the CE community and is believed to gain importance within and beyond CE: Fiedler and Schuppert [76] and Kahrs and Marquardt [50] provided fundamental insight into the identification of hybrid models and extrapolation properties. Furthermore, Kahrs and Marquardt [77] developed methods for determining a valid input domain for hybrid models.

The ML community has identified the need for the incorporation of a priori knowledge for many applications. For instance, researchers in the ML community incorporate prior knowledge as a penalty term in the training and thus enforce physics-informed ANN [78]. A few works also aim to extract physical knowledge from data or data-driven models. Symbolic regression was used to identify physical laws from kinematic data [79]. Interestingly, advanced optimization formulations for symbolic regression have been developed [80] and applied in CE [81].

Overall, CE has a tremendous record in physicochemical modeling and formulating predictive models. Exploiting these capabilities for hybrid modeling is promising to ensure interpretability, extrapolation, reliability, and trust of ML models. At the same time, CE has a strong foothold in (global) optimization of constrained mixed-integer nonlinear problems. Bringing these concepts to the training of hybrid ML methods should reap further profit.

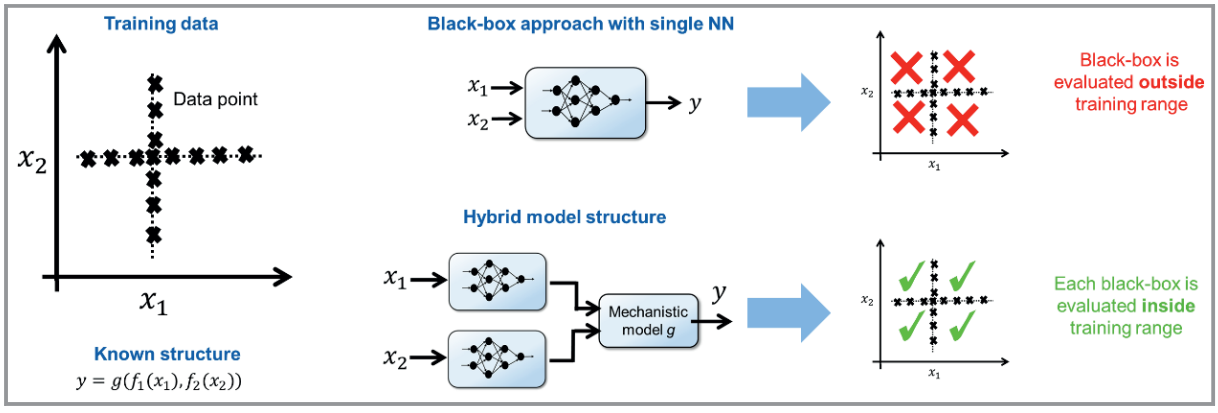


Figure 3. Comparison of hybrid model structure and black-box modeling approach assuming data on a low-dimensional manifold. The dashed lines represent the manifold of the training data points. The mechanistic model g is known a priori. The figure is adapted from a lecture of Andreas Schuppert on hybrid modeling.

Hybrid models that combine data-driven and mechanistic models can avoid extrapolation (c.f. earlier discussion on Fig. 3) and are essential for many CE applications. Furthermore, they improve the explainability of ML models, which matches a current trend in the ML community. In other words, hybrid models are often more explainable compared to a black-box model. For example, intermediate variables in hybrid models commonly have a physical meaning that can facilitate explainability of the predictions. In addition, there is a smooth transition between hybrid modeling and physically motivated ML model architectures. Including physical knowledge into the ML architectures has the potential to enhance generalization and explainability of ML models. For this, CE can build on results from the early 1990s that are today mostly not recognized by the ML community. Most of the previous hybrid modeling efforts can be understood as a top-down approach where a hybrid modeling structure is dictated by the physical and chemical understanding of the system. However, we know from the analogy to expert systems that this top-down approach requires system expertise and leads to high maintenance. Deep learning is successful because it allows for bottom-up (or end-to-end) data utilization. Using the structural information contained in process flowsheets will generate new training schemes for hybrid models and at the same time ensure increased physicality.

3.3 Information and Knowledge Representation

The significant surge of ML applications in social media platforms, online shopping, and video-on-demand services heavily relies on vast amounts of structured data. In CE, however, only a tiny fraction of knowledge and information is accessible for ML methods while the majority is only available in analog or non-standardized digital form. Currently, ML techniques commonly process data from computations, sensors, and measurements. However, molecular data, process flow charts, P&IDs, publications, lab books, etc. are not often accessible to standard ML techniques. This is a major hurdle for finding and exploiting more complex relationships by ML techniques.

Information extraction is the process of (semi-)automated retrieval of structured information from unstructured data sources [82]. For example, natural language processing (NLP) algorithms can recognize entities in unstructured text and extract their relations [83]. Although transformer-based language models have recently demonstrated great advances in NLP [84], automated named entity recognition and relation extraction are still challenging tasks require future research. In addition, figures and tables provide valuable information. Extracting information from tables is domain-independent and there exist multiples tools for this task [85]. However, the extraction of information from figures is often domain-specific and requires joint research efforts [86].

Semantic web technologies connect knowledge and data by using graphs as a unified data model [87]. In particular, knowledge graphs combine data with ontologies, i.e., semantic data models [88]. Currently, there exist only a few chemical engineering ontologies (e.g., ONTOCAPE [89]) and knowledge graphs (e.g., the J-Park Simulator [90]).

In the future, finding new representations for information and knowledge of CE will allow for further analysis, new information and knowledge, and subsequent use, e.g., for optimal decision making. This crosscutting field is hence of great importance to the overall success of ML in CE. In the future, CE data will be extracted from scientific literature and other CE data sources. Moreover, we believe that it will be structured through ontologies and will be saved in knowledge graphs. Using knowledge graph embeddings or other representations allows for automated learning of information [91]. At the same time, the ML methods that work on these specialized knowledge representations need to be tailored to the applications requiring research in both ML and CE. Handling and representing this highly heterogeneous, noisy, and sometimes scarce data is challenging and a key issue that should be addressed in the future; it opens huge potentials in CE but requires both ML know-how and domain-specific insight from CE.

3.4 Heterogeneity of Data

Heterogeneity of data in CE has many sources (e.g., lab books, measurements, property data, molecular simulations, publications, simulation files), and processing heterogeneous data is a major hurdle in CE. Heterogeneity in CE stems, e.g., from (1) multiple scales in time and space (e.g., ms to months in control and scheduling, or nm to m in pore diffusion and pressure swing adsorption), (2) a variety of data sources, which need to be combined to understand chemical processes (e.g., process data, alarms, property data, equipment specifications), (3) highly different data frequencies (e.g., continuous measurement data appears once per ms, while quality data is gathered every other hour or day). All of this is exacerbated by the frequent high dimensionality of data sets [20].

Unsupervised machine learning has emerged in CE for treating high-dimensional problems and perform dimensionality reduction in CE. Recently, an outlier detection algorithm identified strategic molecules for circular supply chains within the "network of organic chemistry" [92, 93] with roughly one million reactions [94]. Breaking down the dimensionality makes huge networks accessible to reaction pathway optimization methods [95–97]. Furthermore, PCA is for example applied [98] to design features from a set of molecular descriptors for solvent selection.

Considering increasingly high-dimensional data sets, manifold learning has become ever more important. Among others, Aimin et al. [99, 100] use manifold learning as the basis for soft-sensor developments for a fermentation

process and a debutanizer column. To tackle highly different data frequencies or data, which is erroneous or incomplete, [20] suggest using semi-supervised learning. It can be employed in case of a mismatch between input and output data, e.g., when a data basis consists of labeled and unlabeled data. A small amount of labeled data can be augmented by larger unlabeled data sets [101]. First preliminary applications in CE use semi-supervised learning to predict missing data for a soft sensor in penicillin production [102].

To describe heterogeneous data, specialized representations of CE information and knowledge are required. For a long time, researchers have designed manual features to describe data. For example, molecules can be described through molecular counts or group contribution methods [103, 104]. However, this manual feature design requires expert knowledge and can lead to a model bias. A promising solution is end-to-end learning, where gradient-based learning is applied on a complete system from the information representation to the output [15]. This has led to breakthrough results in many complex applications including self-driving cars [105] and speech recognition [106]. Recently, molecules and crystals have been represented as graphs and processed by specialized ML algorithms for end-to-end learning [107, 108]. Graph neural networks (GNNs) directly operate on graph structure and have shown promising results for predicting structure-property relationships [109, 110]. Through graph convolutions, GNNs can learn optimal molecular representations and map these representations to the physicochemical properties. As illustrated in Fig. 4, the end-to-end learning approach eliminates the need for manual feature selection. Recent work applied GNNs to predict of quantitative structure-activity and property relationships, e.g., octanol solubility, aqueous solubility, melting point, and toxicity [110, 111]. Further, Xie and Grossman [112] represent crystal structures by a crystal graph that encodes atomic information and bonding interactions for the prediction of target properties. In addition to these applications, GNNs have also been extended to recognize higher-order features from graphs [113]. There have also been some first efforts to represent reaction networks [94, 114] and flowsheets [115, 116] as graphs and apply ML to this data. Venkatasubramanian [7] identified information representation as promising building blocks to further advance the field of CE.

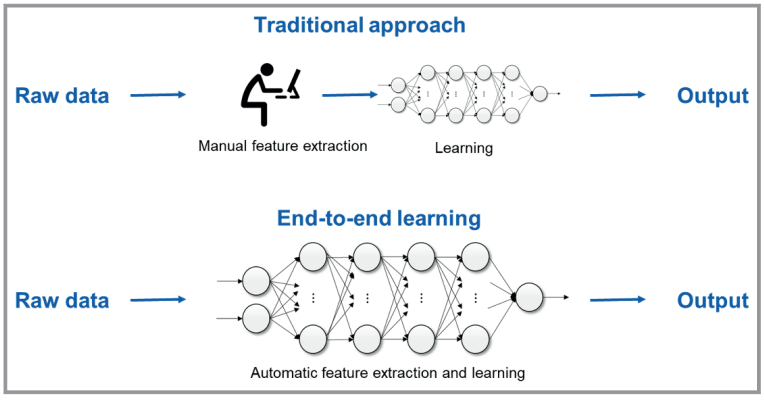


Figure 4. Illustration of the concept of end-to-end learning in comparison to manual feature extraction.

In the future, data of different length and time scales will be combined through ML. For example, simulation and experimental data will be integrated. Beyond this heterogeneity in continuous data, completely different types of data sources will also be integrated. We also expect advanced novel training procedures to construct ML models based on heterogeneous data. For example, recent work predicts experimental procedures from chemical reactions using a transformer language model [117]. Regarding the introduction of data from various scales in space, we expect new modeling paradigms, which automatically rank and subsequently filter the influence of phenomena at multiple scales. The targeted analysis of scientific publications regarding, e.g., thermodynamic information, will allow for the use of historic data in new applications. We envision an automatic identification of similar data sets among a database and subsequent domain adaptation. This necessitates advances in transfer learning and domain adaption. Vice versa, there is a lot of heuristic knowledge in CE on small-scale phenomena, which have tremendous effects at larger scales, e.g., capillary forces and their ramifications for membranes and filters. Formalizing this knowledge and turning it into modeling paradigms for multiscale problems should also be beneficial.

3.5 Safety and Trust in Machine Learning Applications

In CE, a failure could amount to a runaway reaction causing damage to equipment. Safety and trust in ML applications are related to the call for the introduction of physical laws into ML techniques but goes well beyond. It is well-known that the extrapolation capacity of data-driven models over their initial training domain is limited [77, 99, 118]. Thus, the development of models describing the validity domain of data-driven models are desired [77, 99, 118]. However, when training data-driven models on industry data, defining and modeling the validity domain is a major issue [99]. Similar issues can arise when applying GNNs to molecular property prediction [119, 120] or when applying RL to control processes [44, 121]. Overcoming this hurdle is a relevant issue where ML and chemical engineering together can generate considerable added value in terms of research and which would pave the way for new applications.

RL is well-known for its application in game playing, where an agent automatically determines actions that maximize an expected reward

[122]. RL has seen some first and promising applications in chemical engineering for control [44, 121] and scheduling [8]. However, all these initial attempts were purely simulation-based because RL is trial-and-error-based, meaning that experiments can fail. Finally, the field of ML is getting more interested in "explainability" and "interpretability". The usage of black-box models has proven to be a trust issue and more transparent architectures are coming into focus.

In the future, research needs to focus on the safety and trustworthiness of ML methods going beyond just ensuring physicality of ML models. To ensure safety in CE applications, the interpretability of decisions, the provable robustness of models, and the quantification of uncertainty are crucial avenues for future research. In process automation and control, causality- and control-based approaches ensure safety. Integrating these with the promising advances in RL could generate novel methods of use for multiple fields. Similarly, there has been a lot of work on uncertainty quantification both for physicochemical models in CE and for data-driven ML models. These should be brought together for proper descriptions of uncertainty in hybrid models and, of course, for sound design and operation of CE applications under uncertainty.

3.6 Creativity

Creativity is a feature that ML has become quite famous for. Some examples are generating new texts, new sounds, and new images [123]. In CE, a lot of effort using non-ML techniques is currently going into inverse problems, e.g., finding a new catalyst or a novel solvent for a given application [124]. Also, deriving new process structures or new control structures is a desired goal. In ML, matrix completion is a semi-supervised technique that generated a lot of attention by its large-scale application for the "Netflix problem" [125]. Here, predictions are made for non-rated movies based on a large (and sparse) matrix of viewers and ratings. Recently, this technique has been applied to the prediction of activity coefficients for component mixtures, which were never experimentally investigated [126].

Further progress came with generative adversarial networks (GANs), which are deep neural network architectures consisting of two nets competing against one another ("adversarial") [127]. GANs quickly proved to be highly capable at creative tasks such as creating new works of art. Another type of promising generative models are variational autoencoders (VAE) [128], which are also used for data generation (image, sound, text) and missing data imputation.

Overall, CE is a field where novel process designs, products, and materials can currently only be found or discovered by experimental trial and error or by human design. Supporting this with creative techniques from ML might allow for discoveries as of yet unimaginable and research in this direction is hence highly desirable.

In the future, we expect ML methods to solve creative tasks from the thermodynamic phenomena scale up to whole flowsheets and enterprises. Regarding the former, thermodynamic properties for never-measured systems will be inferred by matrix completion techniques.

So far, ML's recent advances on creativity have mainly been appropriately applied on images, texts, and sounds. Transferring these to molecules, control structures, and flowsheets is highly promising as a lot of manual work can be automated and potentially a huge new set of candidate solutions will be found by these techniques. These candidate solutions will also be of great help regarding optimal decision-making. Given the novelty of the applications, further advances in ML methods can be expected regarding training techniques, structuring of data, and metrics for the sensibility of outputs.

4 Conclusion and Outlook

We identified six challenges for interdisciplinary research that will open up new methods for CE and formulate new types of problems for ML: optimal decision making, introducing and enforcing physics in ML, information and knowledge representation, heterogeneity of data, safety and trust in ML applications, and creativity.

The German Research Foundation recently established the Priority Program "Machine Learning in Chemical Engineering: Knowledge meets data." (SPP 2331). The first batch of projects is expected to start in the fall of 2021. With this Priority Program, researchers from ML and CE will jointly work to tackle these emerging challenges. In the meantime, other initiatives have begun researching at the interface of CE and ML. These initiatives promise exciting new research directions in the next few years and will undoubtedly aid in educating a new generation of engineers fluent in methods from both worlds.

The authors gratefully acknowledge the DFG for establishing the Priority Programme SPP 2331 "Machine learning in chemical engineering. Knowledge meets data: Interpretability, Extrapolation, Reliability, Trust". AMS is supported by the TU Delft AI Labs Programme. MK acknowledges support by the Carl-Zeiss Foundation, by the German Research Foundation (DFG) award KL 2698/2-1, and by the Federal Ministry of Science and Education (BMBF) awards 01IS18051A and 031B0770E.

Artur M. Schweidtmann is an assistant professor for chemical engineering at Delft Technical University and co-director of the KDAI Lab, which is part of the TU Delft AI Labs Programme. He received his Master of Science from RWTH University in 2017 and defended his Ph.D. from RWTH in 2021, both in chemical Engineering. During his studies, he spent the academic year 2013/2014 at Carnegie Mellon University as a visiting student via DAAD ISAP program. He performed his Master thesis at the University of Cambridge. His research focuses on the combination of artificial intelligence and chemical engineering.

Erik Esche is a postdoctoral researcher with Prof. Jens-Uwe Repke at TU Berlin. He leads the group's work on development and application of methods for mathematical optimization and machine learning for operation and design of chemical processes. His research focuses on uncertainty in models and measurements and their consequence for the reliable operation of chemical processes.

Asja Fischer is full professor (W3) for mathematics at Ruhr-Universität Bochum, Bochum. She received her Master of Science in Cognitive Science from the University of Osnabrück in 2009 and her Ph.D. in Computers Science in 2014 from University of Copenhagen. Her research focusses on the theory and application of machine learning with a focus on deep learning and probabilistic models.

Marius Kloft is full professor (W3) for computer science at Technische Universität Kaiserslautern. Previously, he was an assistant professor at HU Berlin (2014–2017) and a postdoctoral fellow at Courant Institute of Mathematical Sciences, New York. He earned his PhD at UC Berkeley and TU Berlin (2011). He is interested in theory and algorithms of statistical machine learning, especially unsupervised deep learning, and its applications in chemical process engineering. In 2014, he was awarded the Google Most Influential Papers award.

Jens-Uwe Repke is full professor (W3) for Process Dynamics and Operations at Technische Universität Berlin. He received his Dipl.-Ing. in 1996 and his Dr.-Ing. in 2002, both from the Technische Universität Berlin. From 2010 to 2016 his has been a full professor for Thermal Separation Technologies at TU Bergakademie. His research focuses on optimal process design and operation.

Sebastian Sager is full professor (W3) for algorithmic optimization at Otto-von-Guericke Universität Magdeburg. He received his Diploma (2001), PhD (2006), and habilitation (2012) from Universität Heidelberg, all in mathematics. His research focuses on mixed-integer nonlinear optimization of complex processes and applications in renewable energy, mobility, and clinical decision support.

Alexander Mitsos is a full professor (W3) for chemical engineering at RWTH Aachen University and the director of IEK-10 Energy Systems Engineering at Forschungszentrum Jülich. He received his Dipl.-Ing. from University of Karlsruhe in 1999 and his Ph.D. from MIT in 2006, both in chemical engineering. His research focuses on optimization of energy and chemical systems and development of enabling numerical algorithms. He is the coordinator of SPP 2331.

Abbreviations

ANN	Artificial neural network
ARIMAX	Autoregressive integrated moving average
CE	Chemical engineering
CNN	Convolutional neural network
GNN	Graph neural network
GP	Gaussian process
LSTM	Long-short-term memory
MILP	Mixed-integer linear program
ML	Machine learning
PCA	Principal component analysis
PSE	Process systems engineering
RL	Reinforcement learning
RNN	Recurrent neural network
SVM	Support vector machine

References

[1] A. Kätelhön, R. Meys, S. Deutz, S. Suh, A. Bardow, *Proc. Natl. Acad. Sci.* **2019**, *116 (23)*, 11187–11194.
[2] A. A. Lapkin, in *Handbook of Green Chemistry*, Vol. 12, Wiley-VCH, Weinheim **2018**, 1–16.
[3] A. Mitsos, N. Asprion, C. A. Floudas, M. Bortz, M. Baldea, D. Bonvin, A. Caspari, P. Schäfer, *Comput. Chem. Eng.* **2018**, *113*, 209–221.
[4] J. Artz, T. E. Müller, K. Thenert, J. Kleinekorte, R. Meys, A. Sternberg, A. Bardow, W. Leitner, *Chem. Rev. (Washington, DC, U. S.)* **2017**, *118 (2)*, 434–504.
[5] K. McBride, K. Sundmacher, *Chem. Ing. Tech.* **2019**, *91 (3)*, 228–239. DOI: https://doi.org/10.1002/cite.201800091
[6] A. M. Schweidtmann, A. Mitsos, *J. Optim. Theory Appl.* **2019**, *180 (3)*, 925–948. DOI: https://doi.org/10.1007/s10957-018-1396-0
[7] V. Venkatasubramanian, *AIChE J.* **2019**, *65 (2)*, 466–478.
[8] J. H. Lee, J. Shin, M. J. Realff, *Comput. Chem. Eng.* **2018**, *114*, 111–121.
[9] Y. Bengio, A. C. Courville, P. Vincent, *Representation Learning: A Review and New Perspectives*, arXiv **2012**. https://arxiv.org/abs/1206.5538
[10] G. E. Hinton, S. Osindero, Y.-W. Teh, *Neural Comput.* **2006**, *18 (7)*, 1527–1554.
[11] G. B. Huang, H. Lee, E. Learned-Miller, in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Piscataway, NJ **2012**.
[12] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller, *Playing Atari with Deep Reinforcement Learning*, arXiv **2013**. https://arxiv.org/abs/1312.5602
[13] V. Mnih et al., *Nature* **2015**, *518 (7540)*, 529.
[14] D. Silver et al., *Nature* **2016**, *529 (7587)*, 484.
[15] Y. LeCun, Y. Bengio, G. Hinton, *Nature* **2015**, *521 (7553)*, 436–444.
[16] S. J. Qin, *AIChE J.* **2014**, *60 (9)*, 3092–3100. DOI: https://doi.org/10.1002/aic.14523
[17] C. M. Bishop, *Pattern recognition and machine learning*, 8th ed., Information science and statistics, Springer, New York **2009**.
[18] R. Banares-Alcantara, A. Westerberg, M. Rychener, *Comput. Chem. Eng.* **1985**, *9 (2)*, 127–142.
[19] R. Banares-Alcantara, E. I. Ko, A. W. Westerberg, M. D. Rychener, *Comput. Chem. Eng.* **1988**, *12 (9)*, 923–938. DOI: https://doi.org/10.1016/0098-1354(88)87018-2
[20] Z. Ge, Z. Song, S. X. Ding, B. Huang, *IEEE Access* **2017**, *5*, 20590–20616.
[21] B. De Ketelaere, M. Hubert, E. Schmitt, *J. Qual. Technol* **2015**, *47 (4)*, 318–335. DOI: https://doi.org/10.1080/00224065.2015.11918137
[22] C.-C. Hsu, M.-C. Chen, L.-S. Chen, *Control Eng. Pract.* **2010**, *18 (3)*, 242–253. DOI: https://doi.org/10.1016/j.conengprac.2009.11.002
[23] M. He, C. Yang, X. Wang, W. Gui, L. Wei, *Miner. Eng.* **2013**, *53*, 203–212. DOI: https://doi.org/10.1016/j.mineng.2013.08.011
[24] M. Liukkonen, I. Laakso, Y. Hiltunen, *Environ. Model. Software* **2013**, *48*, 193–201. DOI: https://doi.org/10.1016/j.envsoft.2013.07.005
[25] J. Yu, *J. Process Control* **2012**, *22 (4)*, 778–788. DOI: https://doi.org/10.1016/j.jprocont.2012.02.012
[26] A. Prieto-Moreno, O. Llanes-Santiago, E. GarcÃ-a-Moreno, *J. Process Control* **2015**, *33*, 14–24. DOI: https://doi.org/10.1016/j.jprocont.2015.06.003
[27] M. af Pimentel, D. A. Clifton, L. Clifton, L. Tarassenko, *Signal Process.* **2014**, *99*, 215–249.
[28] L. H. Chiang, R. J. Pell, M. B. Seasholtz, *J. Process Control* **2003**, *13 (5)*, 437–449. DOI: https://doi.org/10.1016/S0959-1524(02)00068-9
[29] M. Yao, H. Wang, W. Xu, *J. Process Control* **2014**, *24 (7)*, 1085–1097. DOI: https://doi.org/10.1016/j.jprocont.2014.05.015
[30] Z. Lv, X. Yan, Q. Jiang, *Chemom. Intell. Lab. Syst.* **2014**, *137*, 128–139. DOI: https://doi.org/10.1016/j.chemolab.2014.06.010
[31] J. V. Kresta, T. E. Marlin, J. F. MacGregor, *Comput. Chem. Eng.* **1994**, *18 (7)*, 597–611. DOI: https://doi.org/10.1016/0098-1354(93)E0006-U
[32] J. Liu, *J. Process Control* **2014**, *24 (7)*, 1046–1056. DOI: https://doi.org/10.1016/j.jprocont.2014.05.014
[33] S. S. Kolluri, I. J. Esfahani, P. S. N. Garikiparthy, C. Yoo, *Korean J. Chem. Eng.* **2015**, *32 (8)*, 1486–1497. DOI: https://doi.org/10.1007/s11814-014-0356-0
[34] Z. Ge, F. Gao, Z. Song, *Chemom. Intell. Lab. Syst.* **2011**, *105 (1)*, 91–105. DOI: https://doi.org/10.1016/j.chemolab.2010.11.004
[35] Y. Yang, F. Gao, *Polym. Eng. Sci.* **2006**, *46 (4)*, 540–548.
[36] P. Jain, I. Rahman, B. D. Kulkarni, *Chem. Eng. Res. Des.* **2007**, *85 (2)*, 283–287. DOI: https://doi.org/10.1205/cherd05026
[37] J. C. B. Gonzaga, L. A. C. Meleiro, C. Kiang, R. Maciel-Filho, *Comput. Chem. Eng.* **2009**, *33 (1)*, 43–49. DOI: https://doi.org/10.1016/j.compchemeng.2008.05.019

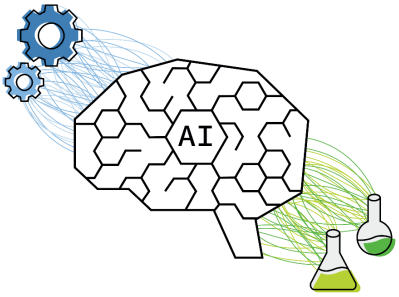
[38] Z. Ge, T. Chen, Z. Song, *Control Eng. Pract.* **2011**, *19 (5)*, 423–432. DOI: https://doi.org/10.1016/j.conengprac.2011.01.002
[39] C. E. Rasmussen, in *Advanced Lectures on Machine Learning* (Eds: O. Bousquet, U. von Luxburg, G. Rätsch) Lecture Notes in Computer Science, vol 3176, Springer, Berlin **2004**.
[40] C. K. I. Williams, C. E. Rasmussen, in *Advances in neural information processing systems* MIT Press, Cambridge, MA **1996**.
[41] H. Kaneko, M. Arakawa, K. Funatsu, *Comput. Chem. Eng.* **2011**, *35 (6)*, 1135–1142. DOI: https://doi.org/10.1016/j.compchemeng.2010.09.003
[42] R. Rico-Martınez, R. A. Adomaitis, I. G. Kevrekidis, *Comput. Chem. Eng.* **2000**, *24 (11)*, 2417–2433.
[43] K. J. Keesman, *System Identification*, Advanced Textbooks in Control and Signal Processing, Springer-Verlag, London **2011**.
[44] P. Petsagkourakis, I. O. Sandoval, E. Bradford, D. Zhang, E. A. Del Rio-Chanona, *Comput. Chem. Eng.* **2020**, *133*, 106649.
[45] S. Hochreiter, J. Schmidhuber, *Neural Comput.* **1997**, *9 (8)*, 1735–1780.
[46] A. Mitsos, A. Fischer, M. Kloft, J.-U. Repke, S. Sager, *Priority Programme Machine Learning in Chemical Engineering. Knowledge Meets Data: Interpretability, Extrapolation, Reliability, Trust (SPP 2331)* **2020**.
[47] S. A. Papoulias, I. E. Grossmann, *Comput. Chem. Eng.* **1983**, *7 (6)*, 695–706.
[48] K. Hornik, M. Stinchcombe, H. White, *Neural Network* **1989**, *2 (5)*, 359–366. DOI: https://doi.org/10.1016/0893-6080(89)90020-8
[49] J. D. Smith, A. A. Neto, S. Cremaschi, D. W. Crunkleton, *Ind. Eng. Chem. Res.* **2013**, *52 (22)*, 7181–7188. DOI: https://doi.org/10.1021/ie302478d
[50] O. Kahrs, W. Marquardt, *Comput. Chem. Eng.* **2008**, *32 (4–5)*, 694–705.
[51] C. A. Henao, C. T. Maravelias, *AIChE J.* **2011**, *57 (5)*, 1216–1232. DOI: https://doi.org/10.1002/aic.12341
[52] I. Fahmi, S. Cremaschi, *Comput. Chem. Eng.* **2012**, *46*, 105–123. DOI: https://doi.org/10.1016/j.compchemeng.2012.06.006
[53] C. Nentwich, S. Engell, in *2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Piscataway, NJ **2016**.
[54] F. Boukouvala, M. F. Hasan, C. A. Floudas, *J. Global Optim.* **2017**, *67 (1–2)*, 3–42.
[55] T. Keßler, C. Kunde, K. McBride, N. Mertens, D. Michaels, K. Sundmacher, A. Kienle, *Chem. Eng. Sci.* **2019**, *197*, 235–245. DOI: https://doi.org/10.1016/j.ces.2018.12.002
[56] Z. T. Wilson, N. V. Sahinidis, *Comput. Chem. Eng.* **2017**, *106*, 785–795. DOI: https://doi.org/10.1016/j.compchemeng.2017.02.010
[57] Q. Zhang, I. E. Grossmann, A. Sundaramoorthy, J. M. Pinto, *Optim. Eng.* **2016**, *17 (2)*, 289–332. DOI: https://doi.org/10.1007/s11081-015-9288-8
[58] M. Mistry, D. Letsios, G. Krennrich, R. M. Lee, R. Misener, *Mixed-Integer Convex Nonlinear Optimization with Gradient-Boosted Trees Embedded*, arXiv **2019**. https://arxiv.org/abs/1803.00952
[59] B. Grimstad, B. R. Knudsen, *J. Global Optim.* **2020**, 1–32.
[60] B. Grimstad, A. Sandnes, *J. Global Optim.* **2016**, *65 (3)*, 401–439.
[61] B. Grimstad, H. Andersson, *Comput. Chem. Eng.* **2019**, *131*, 106580.
[62] R. Anderson, J. Huchette, W. Ma, C. Tjandraatmadja, J. P. Vielma, *Math. Program.* **2020**, 1–37.
[63] J. Katz, I. Pappas, S. Avraamidou, E. N. Pistikopoulos, *Comput. Chem. Eng.* **2020**, 106801.
[64] A. Thebelt, J. Kronqvist, R. M. Lee, N. Sudermann-Merx, R. Misener, *Comput.-Aided Chem. Eng.* **2020**, *48*, 1981–1986.
[65] A. Thebelt, J. Kronqvist, M. Mistry, R. M. Lee, N. Sudermann-Merx, R. Misener, *ENTMOOT: A Framework for Optimization over Ensemble Tree Models*, arXiv **2020**. https://arxiv.org/abs/2003.04774
[66] D. C. Psichogios, L. H. Ungar, *AIChE J.* **1992**, *38 (10)*, 1499–1511. DOI: https://doi.org/10.1002/aic.690381003
[67] A. A. Schuppert, in *Equadiff 99: (In 2 Volumes)*, World Scientific, Singapore **2000**.
[68] M. von Stosch, R. Oliveira, J. Peres, S. F. de Azevedo, *Comput. Chem. Eng.* **2014**, *60*, 86–101. DOI: https://doi.org/10.1016/j.compchemeng.2013.08.008
[69] H. A. Te Braake, H. J. van Can, H. B. Verbruggen, *Eng. Appl. Artif. Intell.* **1998**, *11 (4)*, 507–515.
[70] D. Rall, A. M. Schweidtmann, M. Kruse, E. Evdochenko, A. Mitsos, M. Wessling, *J. Membr. Sci.* **2020**, 118208.
[71] M. Dors, R. Simutis, A. Lübbert, in *Biosensor and Chemical Sensor Technology*, ACS Publications, Washington, DC **1995**.
[72] G. Mogk, T. Mrziglod, A. Schuppert, *Comput.-Aided Chem. Eng.* **2002**, *10*, 931–936.
[73] M. von Stosch et al., *Biotechnol. J.* **2014**, *9 (6)*, 719–726.
[74] P. Schäfer, A. Caspari, K. Kleinhans, A. Mhamdi, A. Mitsos, *AIChE J.* **2019**, *65 (5)*, e16568. DOI: https://doi.org/10.1002/aic.16568
[75] P. Schäfer, A. Caspari, A. M. Schweidtmann, Y. Vaupel, A. Mhamdi, A. Mitsos, *Chem. Ing. Tech.* **2020**, *92 (12)*, 1910–1920. DOI: https://doi.org/10.1002/cite.202000048
[76] B. Fiedler, A. Schuppert, *J. Inst. Math. Its Appl.* **2008**, *73 (3)*, 449–476.
[77] O. Kahrs, W. Marquardt, *Chem. Eng. Process* **2007**, *46 (11)*, 1054–1066. DOI: https://doi.org/10.1016/j.cep.2007.02.031
[78] M. Raissi, P. Perdikaris, G. E. Karniadakis, *J. Comput. Phys.* **2019**, *378*, 686–707.
[79] M. Schmidt, H. Lipson, *Science* **2009**, *324 (5923)*, 81–85.
[80] A. Cozad, N. V. Sahinidis, *Math. Program.* **2018**, *170 (1)*, 97–119.
[81] P. Neumann, L. Cao, D. Russo, V. S. Vassiliadis, A. A. Lapkin, *Chem. Eng. Trans.* **2020**, *387*, 123412.
[82] S. Sarawagi, *FNT in Databases* **2007**, *1 (3)*, 261–377. DOI: https://doi.org/10.1561/1900000003
[83] C. Giuliano, A. Lavelli, L. Romano, *ACM Trans. Speech Lang. Process.* **2007**, *5 (1)*, 1–26. DOI: https://doi.org/10.1145/1322391.1322393
[84] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is All you Need, in *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (Eds: I. Guyon et al.), **2017**.
[85] D. Pinto, A. McCallum, X. Wei, W. B. Croft, in *Proc. of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval – SIGIR '03* (Eds: C. Clarke et al.), ACM Press, New York **2003**.
[86] N. Siegel, N. Lourie, R. Power, W. Ammar, in *Proc. of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, Association for Computing Machinery, New York **2018**, 223–232. DOI: https://doi.org/10.1145/3197026.3197040
[87] P. Hitzler, M. Krotzsch, S. Rudolph, *Foundations of Semantic Web Technologies*, Chapman and Hall/CRC, London/Boca Raton, FL **2009**.
[88] A. Hogan, E. Blomqvist, M. Cochez, C. D'amato, G. de Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, A.-C. N. Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, A. Zimmermann, *ACM Comput. Surv.* **2021**, *54 (4)*, 1–37. DOI: https://doi.org/10.1145/3447772
[89] J. Morbach, A. Yang, W. Marquardt, *Eng. Appl. Artif. Intell.* **2007**, *20 (2)*, 147–161. DOI: https://doi.org/10.1016/j.engappai.2006.06.010

[90] A. Eibeck, M. Q. Lim, M. Kraft, *Comput. Chem. Eng.* **2019**, *131*, 106586. DOI: https://doi.org/10.1016/j.compchemeng.2019.106586
[91] Y. Lin et al., Learning entity and relation embeddings for knowledge graph completion, in *AAAI'15: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, Association for Computing Machinery, New York **2015**, 2181–2187.
[92] B. A. Grzybowski, K. J. M. Bishop, B. Kowalczyk, C. E. Wilmer, *Nat. Chem.* **2009**, *1 (1)*, 31–36. DOI: https://doi.org/10.1038/nchem.136
[93] M. Fialkowski, K. J. M. Bishop, V. A. Chubukov, C. J. Campbell, B. A. Grzybowski, *Angew. Chem.* **2005**, *117 (44)*, 7429–7435. DOI: https://doi.org/10.1002/ange.200502272
[94] J. M. Weber, P. Lió, A. A. Lapkin, *React. Chem. Eng.* **2019**, *4 (11)*, 1969–1981.
[95] K. Ulonska, M. Skiborowski, A. Mitsos, J. Viell, *AIChE J.* **2016**, *62 (9)*, 3096–3108.
[96] K. Ulonska, A. König, M. Klatt, A. Mitsos, J. Viell, *Ind. Eng. Chem. Res.* **2018**, *57 (20)*, 6980–6991.
[97] J. M. Weber, A. M. Schweidtmann, E. Nolasco, A. A. Lapkin, *Comput.-Aided Chem. Eng.* **2020**, *48*, 1843–1848.
[98] Y. Amar, A. M. Schweidtmann, P. Deutsch, L. Cao, A. Lapkin, *Chem. Sci.* **2019**, *10 (27)*, 6697–6706. DOI: https://doi.org/10.1039/C9SC01844A
[99] A. M. Schweidtmann, J. M. Weber, C. Wende, L. Netze, A. Mitsos, *Optim. Eng.*, in press. DOI: https://doi.org/10.1007/s11081-021-09608-0
[100] M. Aimin, L. Peng, Y. Lingjian, *Chemom. Intell. Lab. Syst.* **2015**, *147*, 86–94. DOI: https://doi.org/10.1016/j.chemolab.2015.07.012
[101] X. Zhu, A. B. Goldberg, *Introduction to semi-supervised learning*, Morgan & Claypool, San Rafael, CA **2009**.
[102] J. Ji, H. Wang, K. Chen, Y. Liu, N. Zhang, J. Yan, *J. Taiwan Inst. Chem. Eng.* **2012**, *43 (1)*, 67–76. DOI: https://doi.org/10.1016/j.jtice.2011.06.002
[103] L. Constantinou, R. Gani, *AIChE J.* **1994**, *40 (10)*, 1697–1710.
[104] A. Fredenslund, *Vapor-liquid equilibria using UNIFAC: a group-contribution method*, Elsevier, Amsterdam **2012**.
[105] M. Bojarski et al., *End to End Learning for Self-Driving Cars*, arXiv **2016**. https://arxiv.org/abs/1604.07316
[106] D. Amodei et al., *International conference on machine learning*, New York, June **2016**.
[107] T. Gärtner, P. Flach, S. Wrobel, in *Learning Theory and Kernel Machines* (Eds: B. Schölkopf, M. K. Warmuth), Springer, Heidelberg **2003**.
[108] D. Oglic, S. A. Oatley, S. J. F. Macdonald, T. Mcinally, R. Garnett, J. D. Hirst, T. Gärtner, *Mol. Inform.* **2018**, *37 (1–2)*, 1700130. DOI: https://doi.org/10.1002/minf.201700130
[109] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, R. P. Adams, in *Advances in Neural Information Processing Systems 28 (NIPS 2015)* (Eds: C. Cortes et al.), MIT Press, Cambridge, MA **2015**.
[110] A. M. Schweidtmann, J. G. Rittig, A. König, M. Grohe, A. Mitsos, M. Dahmen, *Energy Fuels* **2020**, *34 (9)*, 11395–11407. DOI: https://doi.org/10.1021/acs.energyfuels.0c01533
[111] C. W. Coley, R. Barzilay, W. H. Green, T. S. Jaakkola, K. F. Jensen, *J. Chem. Inf. Model.* **2017**, *57 (8)*, 1757–1772. DOI: https://doi.org/10.1021/acs.jcim.6b00601
[112] T. Xie, J. C. Grossman, *Phys. Rev. Lett.* **2018**, *120 (14)*, 145301.
[113] C. Morris, M. Ritzert, M. Fey, W. Hamilton, J. E. Lenssen, G. Rattan, M. Grohe, in *Proc. of the 33rd AAAI Conference on Artificial Intelligence*, AAAI Press, Palo Alto, CA **2019**, 4602–4609.
[114] P.-M. Jacob, A. Lapkin, *React. Chem. Eng.* **2018**, *3 (1)*, 102–118.
[115] T. Zhang, N. V. Sahinidis, J. J. Siirola, *AIChE J.* **2019**, *65 (2)*, 592–603.
[116] L. d'Anterroches, R. Gani, *Fluid Phase Equilib.* **2005**, *228–229*, 141–146. DOI: https://doi.org/10.1016/j.fluid.2004.08.018
[117] A. C. Vaucher, P. Schwaller, J. Geluykens, V. H. Nair, A. Iuliano, T. Laino, *Nat. Commun.* **2021**, *12 (1)*, 2573. DOI: https://doi.org/10.1038/s41467-021-22951-1
[118] P. Courrieu, *Neural Networks* **1994**, *7 (1)*, 169–174.
[119] M. Grohe, in *Proc. of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems* (Eds: D. Suciu, Y. Tao, Z. Wei), Association for Computing Machinery, New York **2020**, 06142020.
[120] L. Hirschfeld, K. Swanson, K. Yang, R. Barzilay, C. W. Coley, *J. Chem. Inf. Model.* **2020**, *60 (8)*, 3770–3780. DOI: https://doi.org/10.1021/acs.jcim.0c00502
[121] J. H. Lee, W. Wong, *J. Process Control* **2010**, *20 (9)*, 1038–1048.
[122] Richard S. Sutton, Andrew G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA **2018**.
[123] L. A. Gatys, A. S. Ecker, M. Bethge, *A Neural Algorithm of Artistic Style*, arXiv **2015**. https://arxiv.org/abs/1508.06576
[124] A. Bardow, K. Steur, J. Gross, *Ind. Eng. Chem. Res.* **2010**, *49 (6)*, 2834–2840.
[125] E. J. Candès, B. Recht, *Exact Matrix Completion via Convex Optimization*, arXiv **2008**. https://arxiv.org/abs/0805.4471
[126] F. Jirasek, R. A. S. Alves, J. Damay, R. A. Vandermeulen, R. Bamler, M. Bortz, S. Mandt, M. Kloft, H. Hasse, *J. Phys. Chem. Lett* **2020**, *11 (3)*, 981–985.
[127] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, in *Advances in neural information processing systems* (Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, K. Q. Weinberger) **2014**.
[128] D. P. Kingma, M. Welling, *An Introduction to Variational Autoencoders*, arXiv **2019**. https://arxiv.org/abs/1906.02691

DOI: 10.1002/cite.202100083

Machine Learning in Chemical Engineering: A Perspective

Artur M. Schweidtmann, Erik Esche, Asja Fischer, Marius Kloft, Jens-Uwe Repke, Sebastian Sager, Alexander Mitsos*



Review Article: Recent breakthroughs in machine learning provide unique opportunities for chemical engineering, but only joint interdisciplinary research will unfold the full potential of machine learning in chemical engineering. We identify six challenges of interdisciplinary research that will open up new methods for chemical engineering and formulate new types of problems for ML.

