Research
AI in Chemical Engineering—Review

Transparency: The Missing Link to Boosting AI Transformations in Chemical Engineering

Yue Yuan [a,b], Donovan Chaffart [c], Tao Wu [d], Jesse Zhu [b,c,d,*]

[a] *Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo 315200, China*
[b] *Department of Chemical and Biochemical Engineering, The University of Western Ontario, London, ON N6G 5B8, Canada*
[c] *College of Engineering, Eastern Institute of Technology, Ningbo 315200, China*
[d] *China Beacons Institute, University of Nottingham Ningbo China, Ningbo 315100, China*

A B S T R A C T

The issue of opacity within data-driven artificial intelligence (AI) algorithms has become an impediment to these algorithms' extensive utilization, especially within sensitive domains concerning health, safety, and high profitability, such as chemical engineering (CE). In order to promote reliable AI utilization in CE, this review discusses the concept of transparency within AI utilizations, which is defined based on both explainable AI (XAI) concepts and key features from within the CE field. This review also highlights the requirements of reliable AI from the aspects of causality (i.e., the correlations between the predictions and inputs of an AI), explainability (i.e., the operational rationales of the workflows), and informativeness (i.e., the mechanistic insights of the investigating systems). Related techniques are evaluated together with state-of-the-art applications to highlight the significance of establishing reliable AI applications in CE. Furthermore, a comprehensive transparency analysis case study is provided as an example to enhance understanding. Overall, this work provides a thorough discussion of this subject matter in a way that—for the first time—is particularly geared toward chemical engineers in order to raise awareness of responsible AI utilization. With this vital missing link, AI is anticipated to serve as a novel and powerful tool that can tremendously aid chemical engineers in solving bottleneck challenges in CE.

1. Introduction

Data has rapidly become the backbone of both the market outlook and societal analysis within our time. As a result, modern companies and governments are capitalizing on this valuable asset to create entrepreneurial opportunities and promote regulation development. The optimized utilization of this new resource is attributed to the development of data-driven analysis tools such as data-driven artificial intelligence (AI), which has particularly proved its value through implementation within a myriad of different applications. At the heart of the digital revolution, the value of data-driven AI can be seen through its continuously increasing market avenues. As forecasted by the International Data Corporation (IDC), AI-related worldwide revenue was anticipated to experience an 18.8% growth in 2022 from 341.8 billion USD in 2021, and it is on track to break the 500 billion USD mark by 2024

[1,2]. Although 88% of the current AI market consists of the development of AI software, the market for AI-assisted services is rapidly catching up and is expected to overtake software in the leading AI market role from 2023 onwards [2]. Furthermore, in a report by PricewaterhouseCoopers (PwC) titled "AI Predictions 2021," 86% of over 1000 executives at US companies claimed that AI would be a "mainstream technology" within their company by 2021 [3,4]. These directly claimed interests from executives and CEOs—coupled with the glaring amount of economic benefits—give a strong indication of the inevitable AI transformation-driven changes that are taking place in almost every possible aspect of how the world operates.

Despite AI transformations occurring everywhere, the opacity issue of AI applications partially suppresses their extension to critical domains, particularly those involving safety issues and huge profits. As a result, the revolution brought about by AI utilization is currently distinguishably louder and faster in some particular fields compared with sensitive domains. For example, people are gradually coming to accept options that have been prescreened and customized by AI, such as tailored advertisements on websites

or recommended videos on YouTube. Such applications involve no life-or-death decisions; therefore, recommendations from an AI decision-maker can be readily endorsed. However, in sensitive areas such as clinical applications and chemical engineering (CE) processes, the choices being made are related to health, safety, and tremendous economic value; therefore, the adjudicatures of AI cannot be easily accepted due to a lack of insight into why AI is generating these decisions. As a result, the field of AI as a whole is facing a transparency issue wherein the predictions from AI algorithms are not justifiable, and the rationale behind the decision-making processes is hidden. This is most notable within the currently dominating AI algorithms (e.g., artificial neural networks (ANNs) and deep learning), whose black-box nature further increases the opacity within this field. These models have been gradually taking the lead in AI applications due to their ability to extract hierarchical data representations for accurate predictions of sophisticated systems. However, the tradeoff of using these AI methodologies to acquire ameliorated predictions in highly complex systems is the loss of transparency in decision-making processes [5]. The rising problem of opacity has become a substantial obstacle in the transformative utilization of AI decision-makers in critical domains [6–11]. The lack of AI transparency raises particular concern in the fields of healthcare and safety due to the potential biases and liability issues that stem from the ethical and regulatory aspects of blindly accepting black-box decisions [6]. Nevertheless, momentous interest remains in developing AI decision-makers that can be safely and conscientiously implemented into these kinds of sensitive and vital domains. This interest has led to the exploration and development of explainable AI (XAI) techniques: AI algorithms that can explain and interpret how and why they make their decisions [12]. These techniques have become an immediate necessity for the proliferation of AI within pivotal and crucial areas involving safety and high profitability. These aforementioned critical areas greatly impact our society in profuse different functionalities ranging from clinical operations to energy generation and from medical designs to CE, which serves as the particular focus of this study.

The field of CE spans an extensive scope of vital industries, including petrochemicals, manufacturing, biotechnology, food processing, specialty chemicals, pharmaceuticals, and more. Born from the fusion of mechanical engineering with applied chemistry, CE has always served a heavy duty within industrial practice to satisfy the ever-increasing human livelihood requirements [13]. Accordingly, it is expected that AI transformative utilizations within CE will make significant improvements in our overall living conditions in a vast number of different aspects of CE processes, including design, estimation, control, and monitoring [14]. As CE is an essential domain, it is imperative for the applied AI algorithms in CE to reach high transparency in order to provide reliable predictions that can raise the trust of chemical engineers with a strong sense of responsibility. However, with respect to the explanation and interpretation capabilities of AI decision-makers, the question remains whether the data-driven predictions from AI algorithms are adequate to fulfill the need for transparency in the field of CE. All CE processes are concretely governed by physical and chemical laws that should be considered in both the development of an AI transformation method and the *post-hoc* analysis to reveal the mechanisms controlling the system behaviour. It is thus necessary to merge prior and posterior distributions of physical and chemical laws into AI transformations in order to provide sufficient contributions from the aspect of scientific understanding of CE applications. However, very little special attention has been paid to the transparency of AI transformations within the CE literature. Historically, no special considerations were even given to AI development within CE until 2005 [15], due to the difficulties in the

implementation of sophisticated CE systems. This issue was somewhat remedied following the development of the deep learning method [16], which effectively stimulated AI utilizations in CE [15,17]. This currently dominating machine learning (ML) algorithm enables computers to hierarchically capture and extract data representations and their patterns for convoluted predictions or classifications. However, the "black box" nature of these AI transformation "boosters" requires AI applications in CE to sacrifice transparency in order to adapt to the complexity of the considered systems in this area. In other words, AI utilizations in CE inherently suffer from opacity challenges that limit their applicability within critical CE industrial processes.

Although AI has had a long history of development since it was first proposed by Alan Turing over 70 years ago [18], it is still a relatively new concept to chemical engineers. Consequently, there are various questions that must be dispelled in order to further enhance AI implementation within CE applications. To keep up or even lead the way toward the inevitable AI transformation revolution that is penetrating every aspect of our lives, it is our responsibility as chemical engineers to maintain a clear awareness of what is required of AI applications, particularly within our field of study, in order to avoid the potential stagnation and bottleneck constraints caused by the opacity of AI algorithms. Motivated by the above, this study aims to provide an overview of the current state of transparency in AI applications within the critical domain of CE. Furthermore, this review introduces the essential aspect of transparency to AI transformations within CE in order to fulfill the imperative need for trust and scientific knowledge within this sensitive domain. It should be noted that no comprehensive analyses of AI applications have been previously conducted to elaborate upon the synergetic effects of AI transparency features. This work will thoroughly elaborate on the transparency of AI applications in CE, with a particular focus on the theory-guided aspects of this field, in order to enhance the reliability of these transformative utilizations. Subsequently, this work particularizes the state-of-the-art progress of AI transparency in CE with emphasis on the causality, explainability, and informativeness of these applications. This report subsequently utilizes a thin-film growth case study to provide a comprehensive exploration of AI transparency and to further accentuate the implementation of reliable AI utilizations that can fulfill the demands of CE. Potential research prospects are then discussed to inspire future investigations.

The organization of this paper proceeds as follows. Section 2 presents the identification and detailed interpretation of transparency in AI utilization within CE. Section 3 comprehensively reviews the state-of-the-art progress in three vital aspects of transparency within AI applications in CE—namely, causality, explainability, and informativeness. In Section 4, a systematic case study of an AI transparency analysis is provided in an AI-enhanced multiscale model of thin-film growth. Future perspectives are discussed in Section 5, followed by concluding remarks highlighting the contribution of this review paper in Section 6.

2. From XAI to AI utilization transparency in CE

The nature of work in engineering highlights the value of the sense of responsibility held by practitioners in this area. Accordingly, chemical engineers, who usually bear the burden of ensuring economic benefits and life security in their work, are inherently prudent in accepting non-transparent decision-makers (i.e., modelling algorithms) and non-justifiable predictions. In other words, the transparency of AI transformations in CE determines how far and how fast these utilizations can bloom in this sensitive domain.

Accordingly, XAI is a critical field of emerging AI transformative applications within CE. The original trials of interpreting the decisions made by AI algorithms were arguably conducted by Vaidyanathan and Venkatasubramanian [19] and Swartout and Moore [20] in 1992 and 1993, respectively, and were later deferred due to the focus of AI improvements dramatically shifting to their predictive power [21]. The significant progress of AI development in recent years has led to its extensive application in almost every possible aspect, raising issues regarding the reliability of AI transformations. Fig. 1 displays an analysis of the number of publications related to XAI over the last 31 years on the Web of Science. The keywords "explainable AI" and "explainable artificial intelligence" were used to obtain the presented results, which are displayed according to their year of publication. The shown dramatic increase in the number of studies on XAI—particularly within the last seven years (taking up roughly 99% of the total)—indicates the recent revival of this topic and the necessity to explore the explainability of AI algorithms.

Given the above, what exactly is XAI, and how can it be implemented in CE applications? The general core of this concept refers to transparent AI applications that make their functioning explainable and provide human-understandable justifications for their outputs [5,12,22–24]. However, the concept of XAI swivels between different areas, among diverse audiences, and for various initiatives, since XAI—as a branch of AI—shares similar broad audiences, including computer scientists, medical investigators, social scientists, engineers, and so forth. In particular, the transparency of XAI models is a two-sided estimation that considers both "the ability to explain or to present in understandable terms to a human" [25] and "the degree to which a human can understand the cause of a decision" [26]. Accordingly, explainability and interpretability are two related yet subjectively different cornerstone features of XAI. *Explainability* denotes the active aspect of an AI model in clarifying its internal functions, while *interpretability* refers to the passive aspect of an AI—that is, how well humans are able to comprehend the connections between the input and the predictions [5,21,27]. The transparency of AI transformative utilization within CE is generally rooted in XAI, but it must also align with the characteristics and particular requirements of the field itself. It is not the intent of this work to confound potential practitioners with the debate regarding disunified terminologies. Therefore, this work will specifically elaborate on the transparency of AI utilizations with self-defined terminologies within the area of CE.

The three vital factors that significantly impact the transparency of AI transformations in CE are displayed in Fig. 2, which shows that the evaluation of the transparency of AI utilization can be elaborated upon from the aspects of causality, informativeness, and explainability. Fig. 3 also displays the vital aspects of transparency and their corresponding improvement strategies. Here, *causality* (also referred to as justification) closely links AI predictions to the input features. Accordingly, the predictions of the AI algorithms can be justified by the impacts of their input features, thereby producing understandable explanations of the system behaviour to humans. *Informativeness* implies the ability of AI applications to expatiate CE processes governed by well-known chemical and physical laws in order to advance human understanding of the studied systems. The informative aspect of AI also indicates a model's capability of merging well-known governing laws into its algorithms (i.e., physics-informed or theory-guided AI) in order to obtain physically consistent predictions and improve mechanistic comprehension of the phenomena. Finally, *explainability* represents the transparency of the decision-making processes, depending on how understandable the rationale of the computational flow is—for example, how multilayer structures are operated to improve the ability of a neural network (NN) model to capture the highly nonlinear behaviour of a given system. Thus, XAI transformations within CE should be highlighted by their causality, informativeness, and explainability. More specifically, a transparent AI application can disclose the predictions impacted by the required input features, establish a comprehensive understanding of the phenomenon on the basis of physical and chemical laws, and reveal the rationale for the decision-making process. Enhanced transparency of both the modelling process and the ensuing predictions increases the reliability of AI transformation and ultimately reduces the doubts of chemical engineers. The currently emerging transparency investigations regarding causality, informativeness, and explainability are anticipated to severely affect the extensive AI transformation within CE.

AI applications can be improved with respect to both algorithm transparency and extractive transparency—that is, enhanced transparency from both *a priori* and *a posteriori* analyses, as presented in Fig. 3. *Algorithm transparency* refers to how explainable the modelling process itself is. From this perspective, some AI tools are inherently interpretable, such as the linear related classification method, logistic regression, and decision trees, the latter of which precisely processes a cognitive task in a tree-like structure naturally mimicking the human reasoning process. In regard to the
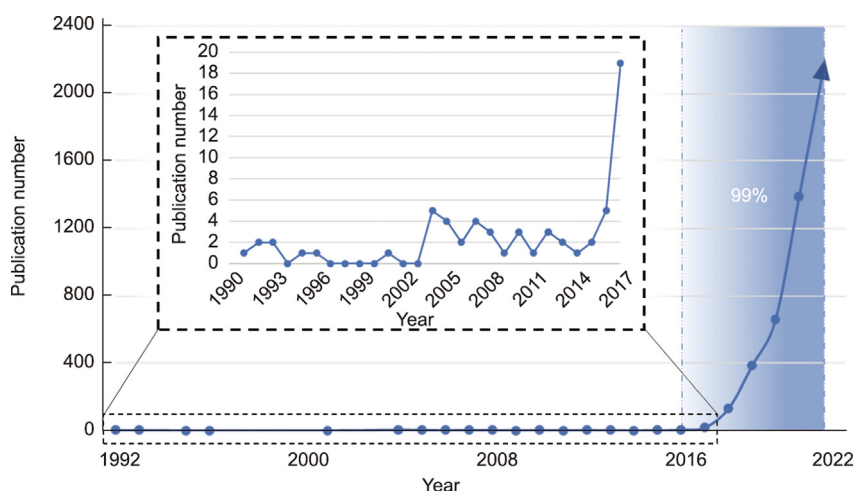

Fig. 1. Number of publications relating to XAI on the Web of Science over the last 31 years, generated using the corresponding categories/keywords of "explainable AI" and "explainable artificial intelligence."
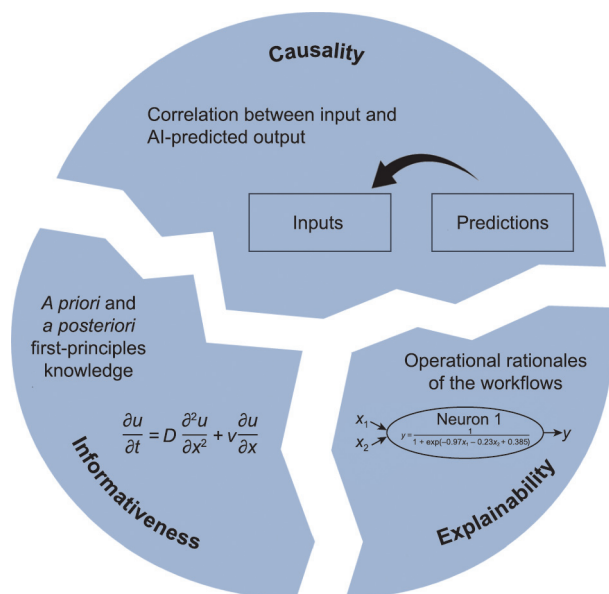
Fig. 2. Illustrations of the causality, informativeness, and explainability transparency concepts of AI transformations in CE.

intuitive interpretability of AI paradigms, symbolic AI—that is, the classical AI of the 1960s–1980s that dominated the expert era through symbolic knowledge and model inference—made significant contributions to AI development with inherent transparency [8,28]. Unlike trendy data-driven AI, symbolic AI can easily explain its logic and reasons by showing which parts of the decision-making processes are evaluated as true or false. Although symbolic AI has not contributed to the current excitement regarding AI development nearly as much as data-driven algorithms (for multiple reasons, including implementation challenges in practical applications), the concept of a hybrid scenario combining symbolic AI and data-driven AI is emerging with the increasing requirements for transparent AI utilization and will be further elaborated on in Section 5 [8,29]. Informativeness also contributes to the overall algorithm transparency. In order to improve the informativeness of AI models, the integration of prior information, such as first-principles knowledge, with ML can stimulate a convergence toward physically consistent predictions. This modelling strategy, which is usually referred to as "physics-informed" or "physics-constrained," has been demonstrated to enhance the transparency of AI utilization in a manner that is especially suitable for expensive data occasions [30,31]. In particular, hybrid modelling—that is, combining first-principles-based models with data-driven algorithms—can contribute to the revelation of the process mechanisms governed by classic chemical and physical laws, thereby

improving the informativeness. Since most CE processes have intrinsic multiscale characteristics, a specific CE phenomenon can be illustrated through an analysis of the electron distribution on the microscopic surface of atoms [32] but can also be explained in terms of macroscopic hydrodynamic behaviour [33]. In hybrid modelling, a multiscale theoretical hybridization strategy can provide a more comprehensive understanding of a system compared with a single-scale theoretical hybridization strategy and can therefore further enhance the informativeness of AI utilizations. However, most current AI applications—such as popular deep learning utilizations—are highly opaque, sacrificing their transparency in order to predict sophisticated system behaviour. Accordingly, *extractive transparency* can be explored before and after the modelling process, the pre-modelling data processing, and the *post-hoc* explainability [5,21,27]. Pre-modelling data processing initially analyses the modelling data to identify the required features for the training and organizes these features into target structures for utilization in algorithms. *Post-hoc* explainability, on the other hand, relates to efforts to explore the understandability of the performed algorithms via a taxonomy of analysis methods such as visualization (e.g., layer-wise relevance propagation, LRP [34]), simplification (e.g., G-REX [35]), feature relevance (e.g., Shapley additive explanations, SHAP [36]), and ensemble learning. In order to thoroughly interpret AI transformations in CE, both the algorithm transparency and the extractive transparency should be considered and evaluated using related technologies and their assemblies.

3. State-of-the-art transparency-considering AI applications in CE

Based on the discussions above, it is apparent that transparent AI utilizations would rapidly promote the progress of the AI revolution by establishing trust in the decision-making algorithms among chemical engineers. Thus, transparency exploration in AI transformations is the vital missing link in developing a reliable decision-making algorithm for critical CE applications. However, the implementation of AI studies within CE is still in its infancy compared with other fields; consequently, there has been very little in the way of AI transparency studies within CE. To effectively spur the corresponding development of this subject matter, this section elaborates upon related state-of-the-art transparency-considering AI applications in CE that analyze transparency in terms of causality, explainability, and informativeness.

3.1. Causality of AI transformative utilization in CE

Among the limited number of transparency-considering AI investigations in CE, the causality of an AI system (i.e., the emerging relationships between a system's input features and the AI
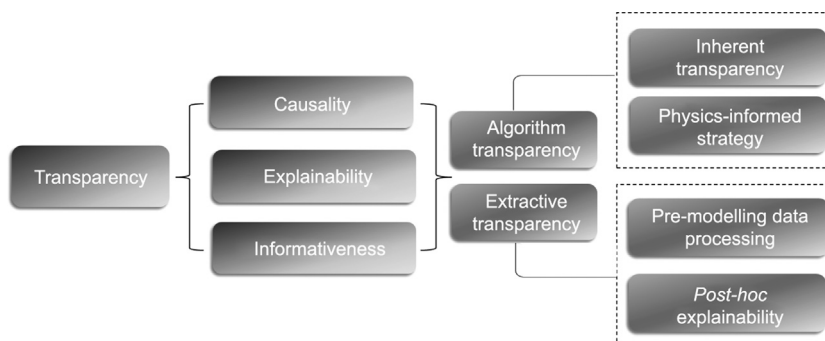

Fig. 3. Transparency of AI transformations in CE and related improvement strategies.

predictions) has increasingly become one of the most prominent areas of study within this field. Although causality-directed AI explorations have only emerged very recently within CE, research in this field has still received some of the greatest attention in CE compared with other aspects of AI transparency [9,37]. Most notably, causality within AI-related CE research has been particularly investigated for systems related to process optimization, control, and monitoring, as the decision-making algorithms in these areas are closely associated with high profitability and safety.

Against this background, Agarwal et al. [38] published an analysis on the causality of an AI algorithm in a CE industrial process that evaluates the classification observability from the input–output data in a straightforward manner while considering the measurement noise in the input data. The observability of the classes indicated the relevance of the inputs for further pruning of the network. Building upon this work, Agarwal et al. [39] further proposed the sequential layer-wise relevance propagation for pruning (SLRPFP) algorithm, which identifies the most highly impactful system inputs and eliminates any non-contributing features within an NN. The researchers subsequently implemented the developed SLRPFP algorithm in two industrial case studies involving the Tennessee Eastman process (TEP) [40,41] and a vaccine manufacturing process. Their work demonstrated that the proposed NN pruning algorithm can effectively avoid data over-fitting issues and can furthermore enhance the computation efficiency of classification models. In addition to contributing toward the causality of these deep learning models, this algorithm contributes to the informativeness of the system through its evaluation of which inputs significantly contribute to the system behaviour. This informative aspect can subsequently be used to estimate both the positive and negative effects of the studied inputs on the profit function.

In another study, Agarwal et al. [23] adopted an LRP method for feature relevance analysis in order to provide causality information for fault detection and diagnosis (FDD) in statistical process control (SPC). Their work implemented the LRP method in their newly proposed FDD algorithm based on a deep supervised autoencoder and a dynamic supervised autoencoder, respectively. The addition of this method suppressed the over-fitting issue present in networks trained on noisy data by iteratively trimming the redundant inputs based on the obtained relevancies from the LRP analysis. These researchers' work revealed that the possible causes of the detected faults could be determined during this process and that the inclusion of LRP improved the FDD test accuracy when compared with previously reported FDD methods.

Aside from feature relevance explainability exploration in FDD, Hale et al. [37] developed inferential sensors based on symbolic regression in order to reduce the evidence of noise and uncertainty in the measured data and improve the robust fault diagnostic process implemented in the heat exchanger fault diagnostic application. Compared with traditional classification methods (e.g., support vector machines (SVMs) and NN methods), utilizing the developed inferential sensors led to greater algorithm transparency due to the algorithm's easily explainable mathematical forms. Thus, these sensors show promise for easier implementation in other systems, compared with traditional AI classification methods. It should be noted that the quality of the classifications reported using the inferential sensors was hardly impacted when compared with those of traditional classification methods, despite the heavily reduced complexity of the sensor-based algorithm.

Based on offline FDD analyses, online process monitoring using AI methods also requires greater algorithm transparency. To address this issue, Bhakte et al. [9] proposed a method to improve the transparency of deep neural networks (DNNs) through the use of SHAP values for variable attribution analysis. These values, which provide a metric to quantify the contributions of system inputs to an AI model's predictions, were implemented into both a simple linear example and the TEP, as well as an online monitoring scenario. The results from this Shapley-based method highlighted which of the numerous measured variables were the most related to the desired outputs, thereby showcasing the method's ability to provide accurate and efficient sample-specific explanations for input–output correlations. Most importantly, this method can provide onsite explanations for understanding real-time operational strategies based on DNN methods.

In addition to FDD analysis, reaction path investigation is another field that has attracted special attention from researchers adopting AI transformative utilization, due to the intensive computing requirements of traditional reaction path methods based on analyzing electronic distributions, such as density functional theory (DFT). Data-driven AI applications have been shown to revolutionarily improve computing efficiencies while adequately maintaining prediction accuracy [42,43]. However, the non-transparent nature of these data-driven AI analyses limits their applications in critical manufacturing processes—particularly those involving sensitive reactions. To address this issue, Kim et al. [44] used LRP analysis to interpret the formation of hazardous reaction products predicted by an NN model in a binary decision task. Furthermore, Kikutsuji et al. [45] conducted local interpretable model-agnostic explanation (LIME) and SHAP analyses to distinguish the contributions of the studied features in DNNs in order to explain the reaction coordinates of alanine dipeptide isomerization. In both studies, the implementation of causality-driven AI techniques provided further insights into the decision-making processes of the AI models, thereby improving the models' overall justifiability for reaction path investigation applications.

Another AI algorithm implementation field that has been critically limited by the lack of transparency pertains to material development in pharmaceuticals for further clinical adoption [46,47]. Particularly within medical applications, it is necessary to have an understanding of the causality of AI implementations from two different perspectives. First of all, it is imperative to have a general understanding of the causal relations between the considered input features and the output predictions. Furthermore, it is necessary to comprehend the human causal understanding, which pertains to the ability of healthcare professionals to understand the relationships of the various factors that influence human health, diseases, and medical outcomes [3,48]. For example, in a case study of histopathology via deep learning analysis and an experienced pathologist, Holzinger et al. [3] emphasized the causality required to achieve "a level of explainable medicine," which featured peer-to-peer explanations to humans to enable their causal understanding. Ward et al. [49] quantified the impacts of the considered drugs on the adverse outcome of acute coronary syndrome (ACS). Their work adopted SHAP and LIME analyses in tree-based models in order to highlight the potential of causality evaluations to accelerate the development of a real-time pharmacovigilance monitoring system. Furthermore, Yang et al. [50] established an efficient 27-layer convolutional NN—a multiscale graph NN (MGraphDTA) based on chemical intuition—for drug–target affinity (DTA) prediction. Their developed model was subsequently used to facilitate drug development, coupled with visual interpretation established by a newly proposed gradient-weighted affinity activation mapping (Grad-AAM) method that labeled the most impactful molecules on DTA. Each of these works emphasized the importance of incorporating AI causality in order to improve the trustworthiness of AI techniques for medical applications. As shown above, causality exploration starts from systems involving critical concerns such as safety and medical utilization and facilitates the application of AI in these fields by justifying AI models' predictions via feature-oriented explanations.

3.2. Explainability of AI transformative utilization in CE

Regarding the explainable rationales behind the AI algorithms themselves (i.e., their explainability), we know of very few studies in CE that have contributed to this area. The reason behind the scarcity of the rationale exploration of AI algorithms in CE is intuitively obvious: The sophistication of CE systems necessitates the use of intricately structured AI algorithms, such as DNNs with highly hierarchical structures (e.g., multiple layers and modified units). As a result, it is intractably difficult to determine the rationales governing the prediction skills of these AI decision-makers. However, the unclear internal mechanisms of these AI algorithms in CE cause reliability issues when the algorithms are applied to critical applications, as discussed in previous sections. Furthermore, the lack of generalizability of these models renders it difficult to avoid possible bias and adversarial examples.

The very first study investigating the internal mechanisms of NNs in CE, to the best of our knowledge, was published in 1992 by Dr. Venkatasubramanian and coworker [19]. This work, which specifically focused on process fault diagnosis in a continuous stirred tank reactor, evaluated the effects on the fault space structures and fault classification performance caused by the number of hidden neurons and the input units in an NN. The increase in the number of hidden neurons can facilitate the separation of the fault space into a greater number of decision regions, improving the ability to capture complicated system information. Single-fault generalization can also benefit from an increased number of hidden units in the NN. Moreover, two-fault generalization can only be improved by an increasing number of input units to deliver more comprehensive information about the diagnosed system to the algorithm. In order to explore this subject further, Dr. Venkatasubramanian's team also conducted the most representative analyses of the rationale investigation of DNNs implemented in CE processes in both classification problems and regression problems in the literature [51,52]. These studies generally demonstrate that complex distributions require more sophisticated NN topologies. In particular, in both classification and regression tasks, deeper networks are capable of transforming the input space into more intricate patterns to simulate relationships of greater complexity and map more intricate distributions. On the other hand, wider networks—that is, networks that have more neurons within a single layer—can combine more simple traits to capture complicated emerging system patterns. Regarding classification problems, an increase in the width of NNs can generally identify a greater number of simple characteristics, thus increasing the generalizability of the NNs to be applied to numerous other tasks. According to the degeneracy analysis of the final layer parameters, countless sets of weights can result in the same accuracy of classification predictions. These scholars' studies also showed that, despite the structures of the NNs, the first-layer nodes will lead to a linear activation of the input space that is successively curved in the later layers. Furthermore, they concluded that similar geometric objects can be required by two separate NNs to perform related but different tasks. This means that the pretrained NN structure can be considered as a reference to initialize the training of another network for a related requirement, which is referred to as transfer learning in the literature [53,54]. Moreover, Dr. Venkatasubramanian's team demonstrated that an increase in the depth of a classification NN is not necessarily beneficial to capturing the actual distributions of data with a possible arbitrary active pattern, and can instead cause adversarial examples that lead to an incorrect diagnosis. The same is true for regression problems: Without a certain width, a deep NN cannot adequately map the input to the output while incurring a sufficiently low loss, especially when the system exhibits a highly nonlinear relation [49]. In other words, the analysis of an NN's workflow indicates that the width of the NN is critical to establishing a high-performance NN. It should be noted that an increase in both the depth and width of the NNs results in sophisticated loss function landscapes, causing difficulties in simulating the regression issues. Moreover, regulation has been shown to be able to reduce over-fitting problems by simplifying the loss of landscapes. In general, the optimal structures of an NN are mainly determined by the generated data and the tasks themselves. Accordingly, for specific CE problems, an optimized performance still requires customized NN structures, which can initialize training based on the discussed impacts of the NN structures or based on pretrained NNs for related tasks.

3.3. Informativeness exploration in CE: Physics-informed and hybrid modelling

Informativeness represents the scientific perspective of engineers, which pursues the understanding of a system from the first principles of the investigated phenomena as *a priori* or *a posteriori* knowledge. However, informativeness also presents a tradeoff with the high performance of the currently dominant AI methods in practical AI utilizations (e.g., advertisements and computer vision). To capture the extremely nonlinear system behaviour, these algorithms primarily rely on massive training datasets with noise, as well as deeply hierarchically and heavily modified operational structures. The mathematical form of these algorithms—and, to a greater extent, the first-principles knowledge—is seldom available or is exceedingly intensive to simulate. Nevertheless, the habit of learning or interpreting systems in mathematical logic is almost an engineer's instinct, developed over the course of engineering training and performing processes for professional operations. This scientific "habit" ensures meticulous planning and general comprehension of the manufacturing processes involved; thus, benefit maximization is achieved while unexpected danger is avoided. Accordingly, the informativeness of AI transformations in CE, achieved by integrating the *a priori* first-principles knowledge and extracting mechanistic insights into the systems, is a vital factor that requires particular attention. To this end, two concepts, ① physics-informed learning and ② hybrid modelling, have raised a certain amount of interest, especially among related practitioners.

Physics-informed learning integrates data and first-principles knowledge ranging from physical laws to empirical, observational mathematical expressions of the phenomena in order to improve ML algorithms [55]. By enforcing the use of *a priori* knowledge in AI models, physics-informed learning processes enforce the theoretical constraints, induced biases, and multi-fidelity observational data patterns, allowing these algorithms to leverage both our prior understanding of the system and observational data patterns to optimize model performance. As a result, these algorithms maintain the key features of analyzing multi-fidelity (and possibly imperfect) data and directing high-dimensional correlations from data-driven AI algorithms, while simultaneously gaining the ability to provide robust, understandable, and physically consistent predictions even in exploration/generalization tasks by exploiting the embedded physical constraints [55–60]. The conceptually simplest integration method to enforce physics in ML is to utilize observational biases that are introduced from data reflecting the underlying physical principles [61,62]. It should be noted that observational biases can only be reinforced with an adequate number of training sets, which may be prohibitive to obtain in engineering applications. A more practical approach is to introduce inductive biases by tailoring NN structures that incorporate prior knowledge for an explicit cognitive task, such as the well-known convolutional NNs, graph NNs, or Gaussian processes [63–65]. However, inductive biases cannot be achieved unless explicit physics and symmetry groups are available with elaborate implementations [55]. Unlike inductive biases, learning biases

are generated by softly reinforcing constraints based on *a priori* understanding of the system through properly penalizing the typical loss function approximations leading convergence toward physically consistent solutions. By means of soft penalty constraints, a partial comprehension of the underlying invariances is incorporated into the model; thus, the implementation of learning bias enforcement tolerates a broad spectrum of first-principles understandings in various domains [55,66,67]. Particularly in CE, the complexity of the studied systems makes it exceptionally challenging to obtain a thorough understanding of the phenomena. The high flexibility and expressive ability of learning-bias-integrated ML algorithms accordingly facilitate numerous studies in CE with the implementation of partial understandings as first-principles constraints, as is the case within the prominent physics-informed neural networks (PINNs). PINNs are especially prevalent in CE as the dominant physics-informed learning technique. Therefore, a further discussion on PINNs is included below.

PINNs are supervised NNs that incorporate the laws of physics in the form of nonlinear partial differential equations (PDEs), including integer-order PDEs, fractional PDEs [68], stochastic PDEs [69], and so forth [70–73]. These physics-driven AI systems have rapidly gained attention as NNs that respect any symmetries, invariances, and conservation principles of physics through the integration of time-dependent and nonlinear PDEs. In these approaches, PDEs are directly incorporated into the loss functions of the NNs as a residual term. The network is then trained to minimize the PDE-based residual over a range of values for each input, so that the PINN can sufficiently predict the PDE output for a given set of inputs. PINNs were first proposed by Karniadakis's group [74] in 2017; the researchers exploited automatic differentiation to represent all of the differential operators, which served as the first trial leading to the recent boost in physics-driven investigations. These early works originally treated PINNs as a complementary method for solving ill-posed fluid problems in which the boundary conditions and/or the initial conditions of the resulting PDEs are unknown [73–75]. Instead, these works achieved PDE solutions through a mesh-free PINN technique that optimizes the loss function of the NN based on the PDE instead of solving the equations directly. Accordingly, PINNs are an effective strategy for establishing an efficient surrogate algorithm for computationally expensive calculations of general nonlinear PDEs. Governed by classic physical constraints such as the Navier–Stokes equations (NSEs), studied systems in CE that suffer from the complexity caused by high-dimensional functions, particularly hydrodynamics, are widely investigated using PINNs. It is well known that investigating hydrodynamics in CE processes via computational fluid dynamics (CFD) is both time and computational-resource consuming. Thus, high-dimensional problems or inverse flow problems are intuitively sophisticated and prohibitively expensive to solve via CFD. Instead, well-trained PINNs can outperform CFD solvers in terms of both accuracy and efficiency, so long as spatiotemporal data is available for both the forward- and inverse-flow ill-posed problems [76]. The flexibility of PINNs in the form of PDE constraints has stimulated PINN applications in a wide range of flow issues, including two-/three-dimensional (2D/3D) flows [73,77], compressible and incompressible flows [56,76], and flow transitions [78]. Although PINN investigations generally focus on flow dynamics, they can be similarly adapted for other applications in CE, such as stiff chemical kinetic issues [30,79], fixed-bed reactors [80], predictive control in consideration of corrupted data [81], flow velocity measurement [82], and material development [83–85]. With regard to ill-posed or inverse problems, PINNs provide more accurate predictions compared with traditional numerical methods, while traditional methods deliver better performance with forward and well-posed issues [55].

The extensive applications of PINNs facilitate the development of software libraries of physics-informed ML. Due to its status as the dominant coding language of ML, most of these libraries are built using Python (e.g., DeepXDE, PyDEns, SimNet) with two exceptions (Neural PDE and ADCME), which are written in Julia. Among them, DeepXDE is a solver that only requires the initial definition of the issues to proceed to the solutions and can handle all computational details automatically. Complex domain geometries can be considered using DeepXDE based on constructive solid geometry. In particular, DeepXDE is highly flexible due to its loosely coupled structure; therefore, it is both an educational and a research tool that can rapidly contribute to the scientific development of ML [70].

In PINNs, the model performance is improved via the hybridization of first-principles-based PDEs and the data-driven NN. From this perspective, the motivation of combining first-principles-based models with data-driven paradigms leads to another appealing informative AI utilization that is particularly prevalent among engineers—hybrid modelling. By integrating mechanistic models with AI methods, the physical understanding of a system can be applied to benefit the algorithms used to model the system in terms of accuracy and robustness [86,87]. The correlations revealed using AI systems with the assistance of mechanistic models can further elucidate the fundamental mechanisms of the investigated phenomena. It is notable that, considering the broad definition of physics-informed learning that enforces first-principles knowledge into data-driven algorithms, hybrid modelling can be considered a physics-informed scenario. However, in the current literature, physics-informed learning generally refers to enforcing physical/chemical understandings within data-driven methods (e.g., enforcing physiochemical equations within the loss function, as is the case for PINNs). On the other hand, in hybrid modelling, the theoretical models and data-driven algorithms can proceed with their calculations both dependently and independently, while maintaining their connections in a parallel or serial manner. Furthermore, the hybridization of data-driven AI methods and theoretical models brings about further advantages in addition to informativeness, such as efficiency improvement, which is another vital factor that attracts much attention to this field. These advantageous properties, along with the currently high interest in this field, can potentially leverage the prosperity of AI utilizations in CE, where they have thus far been deferred. These expedient features will be further discussed below.

Regarding the development of hybrid modelling, the idea of integrating mechanistic models with data-driven algorithms can be traced back to the early 1990s. These early studies aimed at exploiting all available information on studied systems (i.e., mechanistic knowledge, heuristic knowledge, and data with underlying patterns collected from the processes) in a manner that synergically contributed to model establishment for high-quality predictions [32,88–92]. Over the course of 30 years of extensive efforts in hybrid modelling and through its incorporation in a myriad of different applications, the definition of what constitutes a hybrid model has become debatable and ambiguous [14,93]. However, in the context of developing informative AI transformations, hybrid modelling is defined within this work to refer to hybrid algorithms that combine theoretical mathematical paradigms with data-driven AI methods. Some researchers designate the former theoretical expressions as "white boxes," which build mathematical relations from prior first-principles knowledge [86,87]. The latter AI algorithms are accordingly referred to as "black-box models," since they establish the correlations between the model outputs and their inputs in a manner that is hidden to the observer [94,95]. In this context, hybrid models can also be referred to as "gray-box models," due to their integration of both white-box and

black-box models. Physics-informed AI models theoretically belong to the category of gray-box models as well. However, due to their broad interests and conceptual utilization, physics-informed models were previously distinguishably reviewed within this work prior to the present conversation.

As mentioned in the sections above, hybrid modelling can be employed to strategically implement informative AI transformations. Integrating prior physical knowledge of the systems naturally enhances the algorithm's ability to be understood by a human. The operational processes and predictions of the theoretical models within the hybrid algorithms can partially or completely reveal the physical mechanisms of the studied phenomena. Accordingly, the resulting decisions are easier to interpret; thus, hybrid models tend to be more reliable compared with standard AI methods [96]. In addition, a strong hybridization of theoretical and data-driven models can offer direct benefits far beyond informativeness. Although the digital revolution has adequately increased the quantity of available datasets in engineering, CE is still not a "big data" domain, due to its relatively higher expense of data collection compared with other fields of study, such as linguistics [15]. In particular, datasets collected from industrial operations or lab work are generally obtained for systems operating at steady state under normal operating conditions. In other words, these large volumes of datasets collected over a long time span are limited in their variability and thus contain low information content [93]. The inclusion of theoretical modelling can minimize this inherent weakness of AI utilizations in CE by providing substantial low-cost data under a broad range of operating conditions. Consequently, hybrid modelling algorithms are more competent in performing extrapolation and can thus more readily avoid over-fitting [97]. On the other hand, standard theoretical models normally mandate a strong comprehension of a system, including its reaction kinetics, surface revolution, transport phenomena, and so forth. In contrast, the addition of AI methods within hybrid algorithms allows these models to be implemented without a complete understanding of the system. Therefore, hybrid algorithms are easier to implement than standard theoretical models for non-ideal systems for which knowledge of the system behaviour is incomplete. In addition, hybrid models allow for the incorporation of unknown system-oriented parameters correlated into the algorithms to correct the errors of purely theoretical predictions. Furthermore, by replacing computationally intensive calculations with AI algorithms, the efficiencies of hybrid models can be revolutionarily improved [98,99]. AI can also bring high-dimensional features into consideration to detect their synergic impacts in a manner that is much more straightforward than theoretical methods [100,101]. Hybrid modelling is thus prevalent in investigations of CE processes within a non-ideal environment with high complexity.

Hybrid models can be concisely constructed in a serial or parallel order, as presented in Fig. 4. In a serial hybrid model, the data-driven AI models and mechanistic models are both operated sequentially, as illustrated in Fig. 4(a). Accordingly, the outputs of the first model can operate as the inputs to the next model, in a serial fashion. These serial structured scenarios are most helpful

where partial comprehension of the studied system is accessible, which is the case in almost all investigated CE processes. Furthermore, a serial-structured hybrid modelling scenario makes it possible to simulate the inherently sophisticated phenomena of a system without the need to sacrifice the prediction accuracy. As shown in Fig. 4(a), a serial hybrid model can be initiated starting with either the mechanistic module or the data-driven module, depending on the deficiencies of the models, to achieve the objective outputs. These methods are typically referred to as the mechanistic module/data-driven module (MM/DM) and the data-driven module/mechanistic module (DM/MM) serial methods, respectively. It is notable that, in both sequential methods, the data-driven module can provide unknown mathematical representations and process parameters in convoluted environments, thereby substituting for the need for computationally intensive theoretical models and bringing revolutionary efficiency improvement. Meanwhile, phenomenological models undertake the challenge of improving the informativeness of the scenarios [99,102–107]. The MM/DM scenario, in particular, compensates for the expensive data collection in CE. The prior mechanistic module can provide training data under a wide scope of operating conditions, including harsh environments. Endowed with a dataset under harsh conditions in the training of data-driven modules, the hybrid model can effectively diagnose dangerous operating conditions to ensure processing safety. Furthermore, the data-driven and mechanistic models can be operated in parallel and subsequently combined, as illustrated in Fig. 4(b). These parallel scenarios emerge as attractive when the accessible phenomenological mathematical expressions are confined by their prediction power [93]. This confinement can be attributed to the theoretical models' limitations in replicating certain phenomena, such as nonlinearities and dynamic behaviour [86]. The mismatch can be rectified by coupling the results of the data-driven module and the mechanistic module in a parallel manner via Kalman filtering, weighted or not weighted addition, multiplication, and so forth [108]. It should be noted that the structures of the hybrid models are established based on the research objectives; thus, serial and parallel scenarios can be connected, and each module can contain multiple models or backpropagate information to achieve the desired goals [109,110].

Among the discussed hierarchically structured algorithms, AI-assisted multiscale scenarios combining data-driven modules with multiscale theoretical models are known to feature high informativeness. By integrating multiscale theoretical simulations, the intrinsically multiscale phenomena in CE can be considered and illustrated for a comprehensive understanding. Although multiscale theoretical studies are prevalent in investigating CE processes due to their highly informative nature [111–113], the development of AI-assisted multiscale hybrid models has just begun. In studies of hydrodynamics, the ANN–energy minimization multiscale (EMMS) drag scheme was established to provide efficient predictions of fluid dynamics in consideration of local environments by considering the ANN-predicted system heterogeneity in the calculation of drag coefficients by the EMMS model [106,114]. In this work, local gas–particle mixture properties were considered and efficiently connected to the macroscopic fluid dynamics through
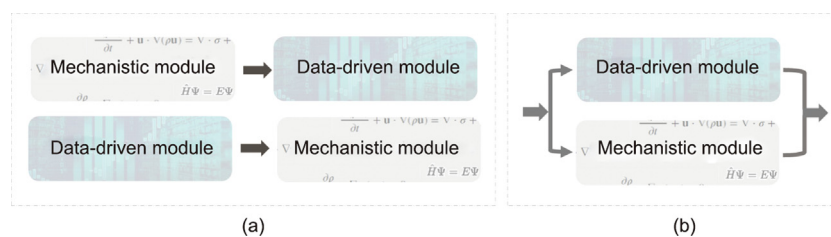

Fig. 4. Hybrid (a) serial and (b) parallel scenarios.

the ANN. It is notable that the above models connected the microscopic environments in a fluid clustering that still presented continuum characteristics. It is particularly difficult to compare microscopic insights that present discrete properties at the electronic or atomic scales with measurable continuum properties for validation. Accordingly, multiscale models that can connect discrete microscopic properties with continuum system features can fill in the gaps in system interpretations between the microscopic studies and the macroscopic investigations. However, there is hardly any investigation in the literature under the umbrella of this scenario, particularly regarding intercommunicating features within the algorithms.

To the best of our knowledge, one of the most significant investigations of hybrid modelling involving data-driven AI assistance and multiscale theoretical modelling, which achieved the aforementioned goal of practical intercommunications, was conducted by Chaffart and Ricardez-Sandoval [115]. In their work, the researchers established a hybrid multiscale model to efficiently model a thin-film deposition process for optimization and control applications. Their model consisted of macroscopic continuum models used to capture the mass, energy, and momentum balances of the gaseous film precursor species within the system, coupled with a microscopic stochastic PDE model to simulate the evolution of the thin-film surface growth on a molecular level. The terms within the stochastic PDE were subject to coefficients whose values were known to vary depending on surface parameters calculated from the macroscale models—that is, the surface precursor concentration and temperature [116]. However, the exact relationship between the PDE coefficients and the surface parameters was unknown. In order to mitigate this issue, the researchers trained shallow feedforward ANNs to predict the stochastic PDE coefficients for given surface parameters. Consequently, the full thin-film growth model developed into a hybrid multiscale model, in which AI learning models established essential communication channels between the macroscopic and microscopic system models. The researchers' proposed hybrid modelling algorithm, therefore, provided one of the earliest examples of using AI to provide intercommunication and translation between microscopic and macroscopic models. However, as the multiscale algorithms that adopt theoretical model assemblies do not in themselves tend to contain multiscale properties, there can be no intercommunication within theoretical simulations, inhibiting their ability to contribute to a thorough understanding of the system.

Motivated by the lack of hybrid multiscale models to connect microscale models with macroscale observations of intercommunicating features, an ideal scenario that could maximize the benefits of AI in CE transformative utilization, known as intelligent intercommunicating multiscale engineering (IIMSE), was proposed, featuring informativeness based on AI-assisted multiscale models connecting microscopic insights with macroscopic behaviour [117]. IIMSE presents highly possible hierarchical structures that combine intercommunicating multiscale theoretical models with data-driven AI algorithms in order to provide a connection between microscopic insights and measurable macroscopic observations. The intercommunicating feature of IIMSE highlights direct information exchange between the theoretical models in a serial order for a comparably complete comprehension of the studied processes. IIMSE can efficiently provide high-quality predictions of the investigated systems while providing a comprehensive understanding to fulfill the scientific goals, allowing it to possibly play a leading role in AI transformations in CE. Moreover, IIMSE provides a solution to connect microscopic insights with measurable macroscopic observations for the validation of the microscopic understanding and the mechanistic interpretation of the observed phenomena. However, this highly potential scenario is still in its infancy, and more attention is needed to facilitate its development. A possible serially constructed scenario is displayed in Fig. 5 as an example of IIMSE for readers' reference. In this scheme, a data-driven ANN connects a microscopic multiscale model to a macroscopic CFD analysis. Through this connection, the mesoscale models, which are difficult to mathematically represent, are substituted by the data-driven model to facilitate the implementation of this whole-scale simulation—that is, a simulation from the microscopic scale to the macroscopic scale—while maintaining reasonable efficiencies in order to predict the macroscopic measurable system performance in a large experimental domain.

The preceding mechanistic module encloses a three-scale intercommunicating theoretical model. This microscopic multiscale model is established by combining quantum chemistry analysis (i.e., DFT), molecular simulations (i.e., kinetic Monte Carlo, kMC), and continuum mass transportation equations inside a particle channel, as shown in Fig. 6. The elementary reaction mechanisms, as well as the relevant energetic features, are calculated from DFT analysis and then transferred to the kMC model as surface events and reaction rates. The kMC analysis subsequently predicts the dynamic surface features, such as the diffusion rates and concentrations, to establish the boundary conditions of the continuum equations in the particle channel. The predictions of the continuum models, such as the coverages of the surface species, can be reversely delivered to the surface model of kMC to impact the surface evolution. This multiscale intercommunicating microscopic model has been partially demonstrated to be efficient in contributing to the comprehensive understanding of a CE process [91,118]. Posterior CFD analysis of the ANN can be initiated from the empirical parameters, such as the overall reaction rates connected to the
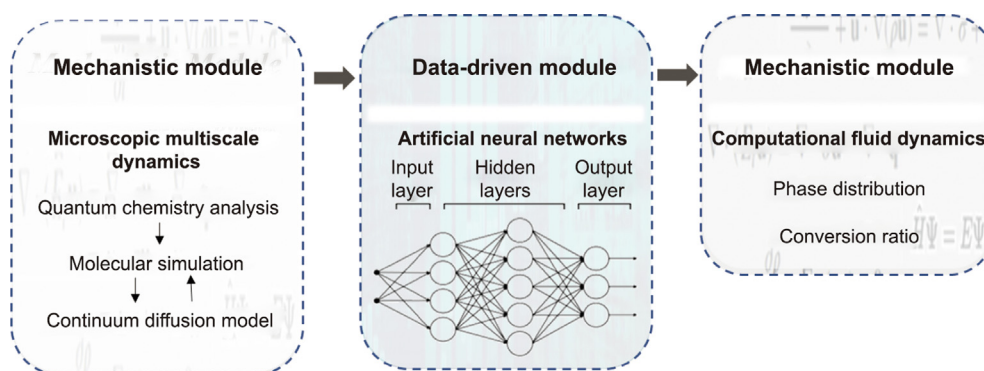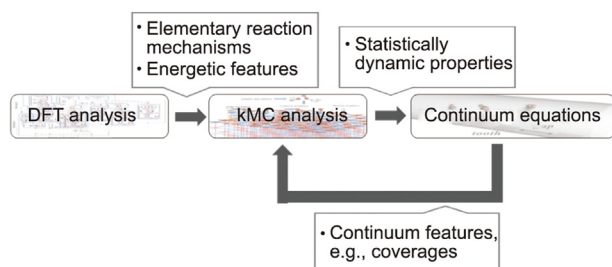

Fig. 5. Example framework of IIMSE modelling.

Fig. 6. Mechanistic module of the IIMSE example.

microscopic states simulated by the microscopic intercommunicating multiscale models. Accordingly, the experimentally measurable system performance, such as the heterogeneous phase distribution and the substance transformation, is predicted by CFD analysis. The example IIMSE scheme has the potential to provide a solution to implement a whole-scale simulation of a CE process that can relate the microscopic states and mechanisms to the measurable system performance. In particular, the deep integration of theoretical models with data-driven algorithms dramatically improves the informativeness of the hybrid model, contributing to the scientific understanding of the investigated phenomena. Furthermore, the connection between the microscopic states and the measurable system features opens up the possibility of justifying the observable system behaviour using microscopic insights.

Informativeness is generally the most studied aspect of the transparency of AI utilization within CE applications. Physics-informed learning and hybrid modelling, both of which integrate first-principles knowledge into algorithm development, have made a major contribution as effective strategies to enhance the informativeness of AI transformative utilization in CE. Physics-informed learning, which leads the convergence toward physically consistent predictions by introducing biases rooted from first-principles comprehension, is a broader concept that encompasses hybrid modelling, which couples theoretical models with data-driven algorithms. Despite some overlaps, hybrid modelling usually generates independent predictions from theoretical models, while physics-informed learning induces theoretical constraints within the data-driven algorithms, based on current publications. Both of these methods contribute toward the informativeness of a system, and thus fulfill the scientific requirements of AI utilization in solving CE issues.

4. A case study of AI transparency exploration

In the previous sections, transparency investigations in AI applications were elaborated on regarding informativeness, causality, and explainability. These advanced analyses contribute to raising trust in critical domains, particularly in the focused field of this manuscript, CE. To further highlight the significance of AI transparency analysis and present an example to stimulate further related explorations in CE, this section presents a case study that provides a comprehensive analysis of transparency involving all three discussed features, as shown in Fig. 7.

The case study considers a simplified version of the hybrid multiscale thin-film deposition model discussed in Section 3 and outlined in Fig. 7. It should be noted that the development of the hybrid multiscale model within the proposed case study was previously established within the literature [115]. This system depicts

a vapor deposition chamber with a substrate placed inside. The chamber is subsequently filled with gaseous precursor species that migrate to the substrate surface and deposit on it to form the thin film. The behaviour of the precursor gas species within the deposition chamber is captured using macroscale continuum equations that describe the precursor species mass and momentum balances according to the macroscopic gas-phase model equations [115]. These equations are used to predict the molar fraction of precursor species at the thin-film surface y_s, which—in addition to the externally controlled substrate temperature T_s—serve as the key variables affecting the thin-film growth. The growth of the thin film is captured on a molecular level by the stochastic PDE shown under the title "Microscopic thin-film model" in Fig. 8. In this expression, h denotes the local height of the film at location x along the substrate and at time t; v, K, and F denote coefficients whose values depend on y_s and T_s; and ζ is a random Gaussian variable with mean 0 and shot noise covariance $\langle \zeta(x,t), \zeta(x',t') \rangle = V\delta(x - x')\delta(t - t')$. It should be noted that each term in the stochastic PDE coincides with a specific molecular-level kinetic event that is expected to occur on the surface during thin-film growth: The term F corresponds to the adsorption of the precursor onto the film surface, whereas the terms $v\frac{\partial^2 h}{\partial x^2}$ and $K\frac{\partial^4 h}{\partial x^4}$ are respectively associated with the desorption and migration of adsorbed surface molecules. The results from the stochastic PDE expression are used to calculate the average film thickness $H(t)$ at time t, which can be subsequently used to calculate the term $R_a - R_d$ (the difference between the rates of adsorption (R_a) and desorption (R_d)) from the macroscopic mass-transport surface boundary condition. Further details of the multiscale model can be found in Ref. [115].

In the above multiscale model, the coefficients v, K, and F and the stochastic covariance V are all dependent on the surface parameters y_s and T_s. However, the exact relationship between the stochastic PDE coefficients and the surface parameters is unknown [115]. In order to bridge the gap between the macro and micro scales, a series of shallow feedforward ANNs are trained to predict the values of v, K, F, and V as a function of the surface inputs y_s and T_s, as illustrated in the "ANN" segment of Fig. 7. To accomplish this objective, the PDE coefficients are determined at discrete values of y_s and T_s by model-fitting the stochastic PDE to a kMC-based surface growth model, as highlighted in Ref. [115]. The generated pairs of input (y_s, T_s) and output (v, K, F, V) coefficients are subsequently used to train individual ANNs for each of the output coefficients via backpropagation using the Levenberg–Marquardt algorithm. Each ANN contains a single hidden layer consisting of 10 neurons for the ANNs predicting v, F, and V, and 15 neurons for the ANN predicting K.

The informativeness of a model is a key criterion that can be used to provide a greater understanding of the underlying physical and chemical processes affecting the system behaviour. As highlighted in Section 3.3, model informativeness can be indicated from both *a priori* and *a posteriori* knowledge. In the proposed hybrid multiscale model, a great deal of the system behaviour is known in advance, due to the incorporation of known physical transport equations into the macroscale gas-phase model. These transport equations integrate key knowledge into the model regarding how the gaseous precursor species behave within the reactor chamber and on the state of the gas phase adjacent to the growing thin-film surface. In addition, due to previous studies, the stochastic PDE used to capture the molecular-level surface film growth is known to provide insights into the physiochemical kinetic processes taking place on the surface [116], despite the lack of a microscale mechanistic model, as highlighted in the paragraphs above. These previous studies have specifically
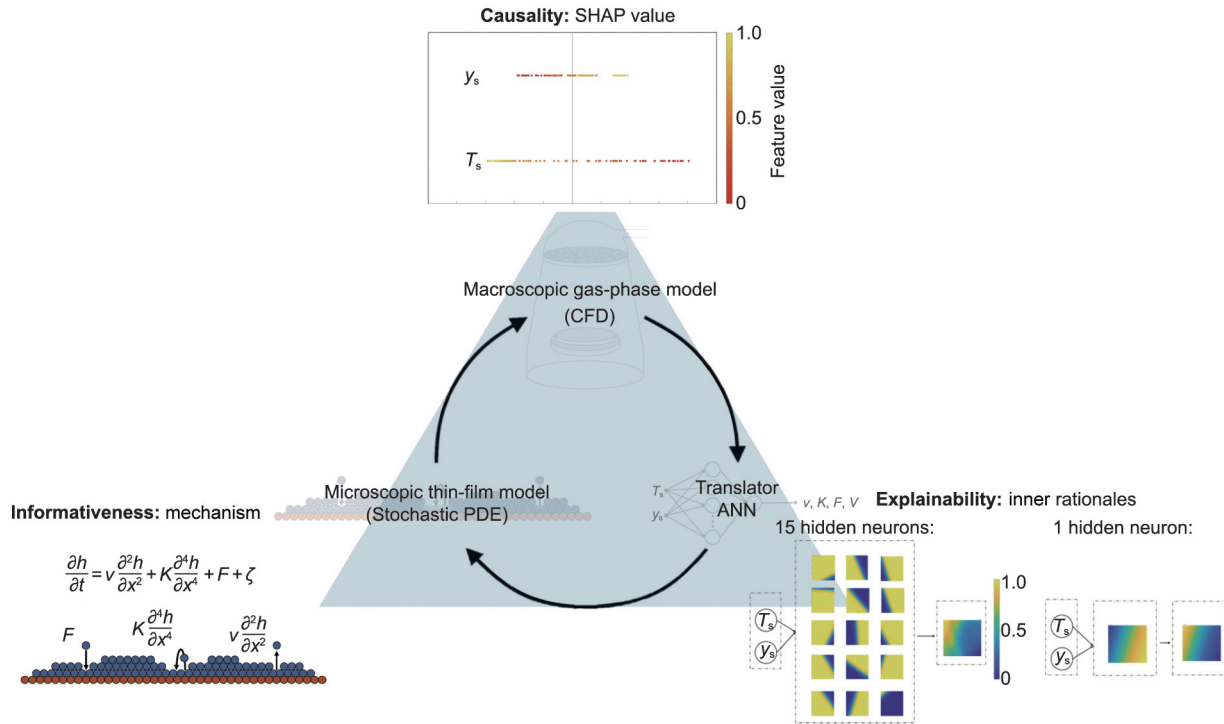
Fig. 7. A comprehensive transparency analysis of a thin-film growth model from the aspects of informativeness, causality, and explainability.

demonstrated that increasing the values of the PDE coefficients F, v, and K results in greater amounts of adsorption, desorption, and migration, respectively, on the film surface.

In addition to the system knowledge determined *a priori*, the fully assembled hybrid multiscale model can be used to gain further mechanistic insights into the inner workings of the processes taking place. In Ref. [115], it was observed that lowering the surface temperature within the hybrid multiscale model promoted higher rates of surface adsorption and resulted in higher surface growth rates and higher surface roughnesses. On the other hand, higher surface temperatures were observed to promote higher rates of desorption and surface migration, and therefore resulted in lower surface roughnesses and smaller growth rates. These results indicate the effects of temperature on each of the kinetic processes taking place and specifically emphasize that the rates of desorption and migration are more significantly impacted than adsorption by the temperature. The hybrid model also demonstrates the effects of changing the concentration of the precursor species within the gas phase on the surface growth behaviour. More specifically, the model predicts that increasing the precursor species fraction results in larger growth rates and thus higher

surface roughness. The opposite was observed when the precursor species fraction was decreased. These results indicate that surface adsorption is most strongly affected by the precursor species concentration; therefore, more adsorption is experienced when the concentration increases. Furthermore, the full hybrid model can be used to provide insights into how to achieve explicit surface specifications by controlling the thin-film growth process. For example, if the aim is to maximize the surface growth rate while producing a surface with minimal roughness, then the model predicts that it will be necessary to keep the temperature low during the early stages of growth, in order to prioritize adsorption, and then ramp up the temperature toward the end to maximize the surface migration events taking place and produce a smooth final film [115].

In comparison with informativeness and explainability, causality indicates the correlations between the AI model inputs and outputs, which would provide efficient guidance for practical operations in particular. In Fig. 9, the SHAP analysis (a prevalent tool for correlation analysis from inputs to outputs) of the ANN
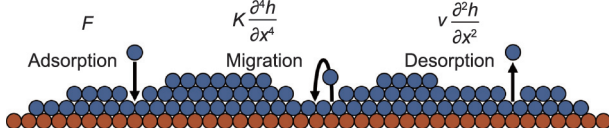

Fig. 8. The informativeness of the case study, as assessed through its inclusion of first-principles models such as the depicted thin-film surface model.
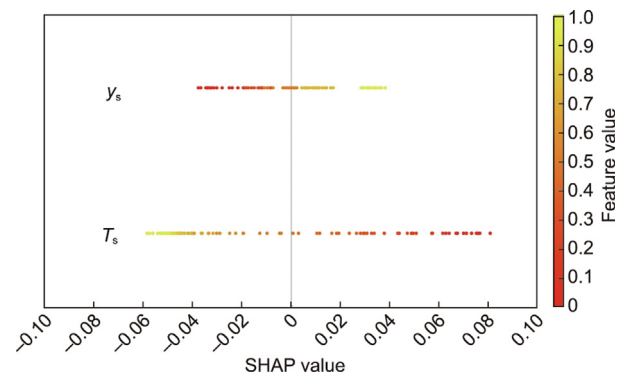

Fig. 9. The causality of the case study, as assessed through SHAP analysis.

used in the explained hybrid model is presented to generate the impacts of the inputs, the precursor fraction, and the temperature on the output coefficient K. The result displayed in Fig. 9 indicates that a high concentration and low temperature on the surface of the thin film lead to an increase in K, which generally suggests a drop in the migration rate, since K is inversely related to the migration rate [116]. This comes from the fact that, within a low range of temperatures, a temperature increase leads to a higher adsorption rate. When the adsorption dominates, it creates a large number of highly unstable surface configurations, resulting in an increase in migration rates as the unstable surface sites diffuse across the surface to form a more stable surface configuration. However, within a higher range of temperatures, the rate of surface migration increases due to the Arrhenius relation between migration rates and temperature. Therefore, there is a competition within the calculation of the migration rate between the Arrhenius behaviour and the effects of the unstable surface sites available on the surface. With regard to the precursor fraction, when the precursor gas concentration is fairly high, the surface is supposed to have a high number of occupied sites. Since the adsorbed molecules are always attempting to move to form a more stable configuration, a more occupied surface caused by an increase in the precursor gas concentration would lead to higher migration rates. However, when the precursor fraction is lower, the adsorption rate for the surface is low; subsequently, there are very few molecules adsorbed onto the surface of the film. Due to the low concentration of the adsorbates, each migration event across the surface has a lower chance of forming a more stable surface configuration. Consequently, the loose surface adsorbates will need to undergo a significantly greater number of surface migration events in order to encounter other loose adsorbates and form a more stable surface configuration, since the stability of a surface molecule increases when it has a larger number of nearest neighbors [116]. As a result, lower surface precursor species fractions result in a greater number of migration events due to the number of migrations that need to occur to form a stable surface configuration.

With regard to the explainability of the established ANN, Fig. 10 compares two ANN structures for the same task. Both have one hidden layer, but one has 15 neurons (as used in the previous thin-film model) and the other has just one neuron. It is clear in the comparison that the original ANN used in the hybrid model with 15 neurons can bring about a highly nonlinearly activated output space, as shown, while the other ANN that only adopts one neuron in the hidden layer can only lead to a linearly activated output space. This result indicates that, even with only one hidden layer, the original ANN can still capture the nonlinear behaviour of the system while using 15 neurons. In the first and only hidden layer, all the neurons in both ANNs present a linearly activated space. However, the combination of 15 linearly activated spaces generates fairly high-quality predictions of the nonlinear system, whereas the shallow NN with one neuron cannot achieve a similar result. This comparison suggests that the width of the hidden layer of the original ANN ensures the capture of many linearly activated details of the system, which leads to the high prediction performance of the trained ANN for the studied system, while maintaining high efficiency.

Overall, the transparency analysis in this case study enhances the mechanistic understanding of the system, reveals the environmental factors impacting the studied dynamics, and illustrates the operational rationale of the ANN. In this way, effective guidance is delivered that contributes to the scientific comprehension, thin-film manufacturing, and further model establishment of the data-driven algorithms.

5. Perspectives on transparent AI

The transparency of AI applications within CE is proposed for the first time in this work; hence, no comprehensive application covering all aspects of transparency was available in the literature *a priori*. Furthermore, AI applications concerning any of the aspects of causality, informativeness, and explainability are currently
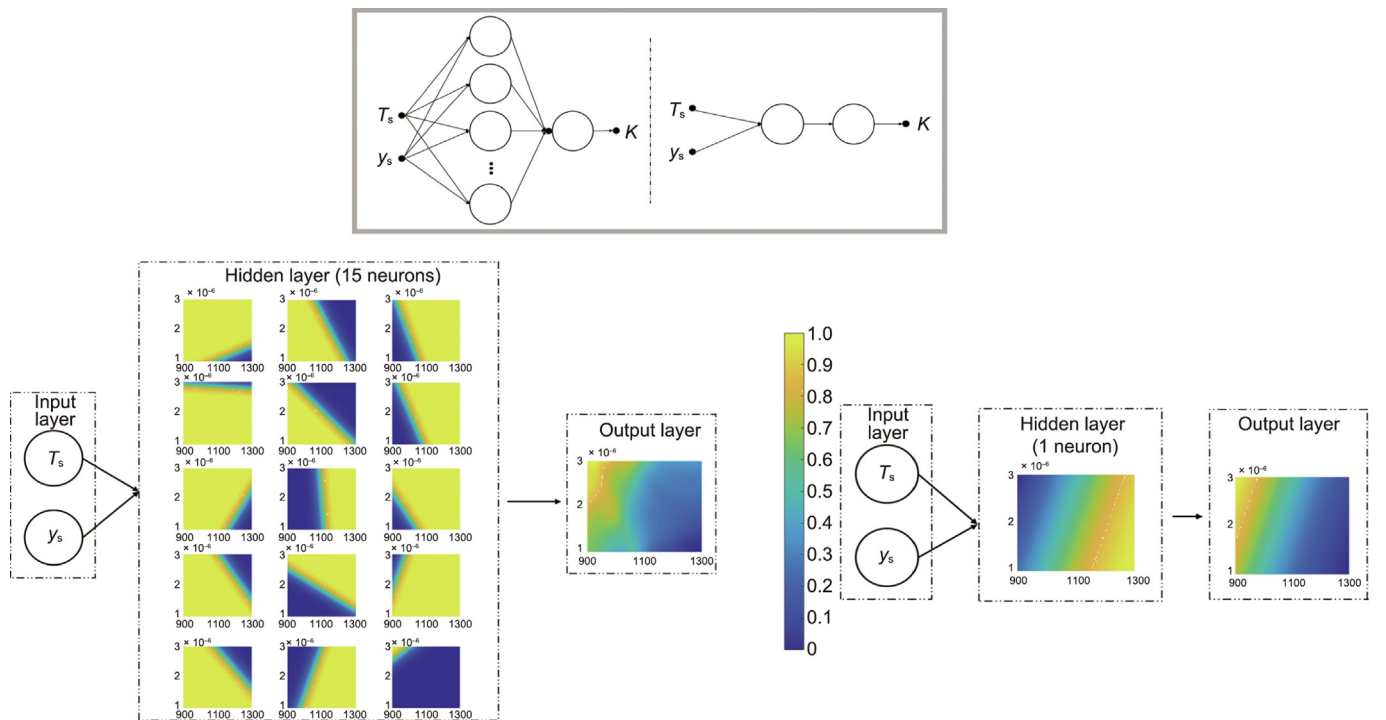
Fig. 10. The explainability of the case study, as assessed through an analysis of the NN structures predicting the stochastic PDE coefficient K using 15 neurons (left) and using a single neuron (right).

limited within the context of CE, as discussed in the previous sections. As a result, there remains a significant need to advance the study and implementation of AI transparency techniques, particularly in the field of CE. This section will accordingly discuss the limitations and future research directions of AI transparency from the viewpoints of causality, informativeness, and explainability.

Causality, as a distinct defining aspect of AI transparency, has piqued the interest of chemical engineers for a relatively long period of time. This justifiability, in turn, makes it possible to optimize the observability of the operating processes in order to balance process controllability with economic benefits. Therefore, there is a great need to expand our current knowledge of AI process causality in order to improve the reliability of AI applications within CE. Model-agnostic *post-hoc* analysis is naturally more accessible compared with inner-rationale exploration and mechanism revelation. Accordingly, the most investigated applications are in post-operational SPC studies regarding feature relevance explainability [9,23]. In other words, it is essential for real-time explanations that dynamically modify onsite strategies to meet safety and efficiency requirements to be initiated in order to facilitate the revolution of intelligent manufacturing toward Industry 4.0 [93]. Accordingly, the future of causality research within AI applications lies in pushing beyond the boundaries of our current techniques to provide dynamic and on-the-fly explanations for AI decision processes in any general CE application, based on the live extracted dependencies of the model outputs on the system inputs.

In addition to the causality of AI transformations, the need for scientific interpretations in CE has led to a flourishing of informativeness development within AI applications. However, despite their intuitive promise of greater AI transparency, many of the proposed hybrid and physics-informed modelling methods have only been partially explored at present and require extensive further study. With the assistance of powerful simulation algorithms and accessible computing recourses, highly hierarchically constructed hybrid models are encouraged in order to achieve high prediction performance and thorough comprehension. In particular, IIMSE, through its combination of intercommunicating multiscale theoretical models and data-driven AI paradigms, holds great promise for playing a leading role, with its high adaptability to CE investigations and strong potential to maximize the benefits of AI to solve bottleneck issues existing in CE. The features of IIMSE that connect microscopic insights to measurable macroscopic behaviour open up possibilities to efficiently predict CE operations from microscopic material design to optimized system operations for specific objectives, such as pharmaceutical development and oxygen carrier exploration. Thus, further efforts should be focused on developing whole-scale IIMSE models to capture different CE applications from the electronic to industrial scale. Moreover, hybrid models that exploit theoretical models to provide datasets with greater information content are required to compensate for the performance of systems under harsh condition, as such datasets are otherwise difficult to obtain for safety reasons. Alongside IIMSE, the hybridization of symbolic AI techniques and data-driven AI algorithms, such as the explainable artificial intelligence framework for mechanistic explanation generation (XAI-MEG) [8], presents another solution to the deficiency of AI utilization in terms of both mechanistic insights and causal explanations, and thus deserves further attention. Likewise, the integration of ANN and fuzzy systems has provided multiple modelling applications that process interpretation capacity, such as fuzzy neural networks (FNNs) and neuro-fuzzy networks. In fact, FNNs and neuro-fuzzy networks have gained attention from researchers within diverse fields since the 1990s. Their inherent strong adaptability, inherited from fuzzy systems, has empowered them with extensive application capacity. Accordingly, more significant contributions from these hybrid models can be expected [119].

Regarding the inner rationales underlying developed AI systems, the scarcity of related studies in the field of AI explainability indicates the particular difficulty of such investigations. Thus, even among the existing analyses on the explainability of DNNs implemented in CE, it cannot be ignored that exploration of the internal mechanisms of AI within CE is extremely limited, compared with the numerous AI-related studies. Furthermore, much of the current focus on AI explainability has only been applied to DNNs thus far. As a result, there is a need for a tremendous endeavour to elucidate the operational mechanisms of critical techniques along with numerous other widely used methods, such as recurrent connections, convolutional filters, reinforcement learning, and more [120–122]. Such research pathways are crucial within the future research of CE AI implementations, in order to advance the understanding of the delineation of decision regions within the operational processes of multilayered AI algorithms and contribute to the ultimate explainability of the predominant deep learning applications. In keeping with this perspective, the prosperity of AI transformative utilizations should be vigorously promoted by illustrating the effects of AI networks' hidden units in order to improve the performance of deep learning paradigms and guide their extensive application in other CE processes. Step-by-step comparisons from simple structures to multilayered systems have already been demonstrated to be efficient within these investigations [51,52]. However, similar work is required to study deep learning assemblies that parallelly exploit inherently transparent proxy models such as SVM or decision trees as an additional route within inner-rationale exploration.

Transparency research within AI applications in CE is a growing field that is in high demand, and whose need is greater than ever. The pressing demand for related research necessitates tremendous efforts to promote the practical use of AI in CE. Consequently, there is still much work that must be conducted in order to fully incorporate transparency implementations within AI applications in CE.

6. Concluding remarks

The tremendous global revenues created by AI transformations indicate the inevitable change brought about by this revolutionary technology. The predominating AI paradigms enable the ameliorated predictions of sophisticated CE processes via hierarchically extracting data representations. However, as a tradeoff for the high prediction performance of AI, these AI transformations experience greater model opacity. Such non-transparent algorithms can subsequently cause ethical and regulatory issues by providing non-justifiable decisions. Particularly in critical domains such as CE, which involve health, safety, and huge economic value, greater transparency of AI applications will directly lead to an increase in the reliability of these decision-makers for humans. Accordingly, transparency is the vital missing link needed to boost AI transformative applications in CE.

Motivated by the scenario described above, this review defined for the first time the vital concept of transparency within AI utilizations in CE, rooted in XAI, with a special focus on the characteristics and requirements of the CE field. This concept of transparency was elaborated upon in terms of the aspects of causality, explainability, and informativeness—that is, the correlation between predictions and the considered input features; operational rationales hindered by decision-making processes; and the mechanistic insights of the investigated systems. These methods, along with their state-of-the-art applications regarding all three essential transparency factors, were discussed and compared in order to highlight the significance of reliable and responsible AI utilization, thereby stimulating related investigations.

Informativeness in particular has recently become a current focus in the transparency exploration of AI applications in CE. This focus stems from the inherent features of the CE field, the comparatively expensive information patterns from datasets, and the necessity to reveal the underlying mechanisms for system comprehension. The hybridization of prior knowledge, including first-principles understanding and theoretical models, and data-driven AI methods provide a solution to enhance the informativeness of these applications. More specifically, extensive investigations in physics-informed and AI-assisted hybrid models highlight their effectiveness in contributing to providing mechanistic insights into systems with sparse datasets.

Causality, on the other hand, has piqued notable attention from chemical engineers, who aim to optimize the observability of system behaviour while balancing the system controllability and economic benefits. The impact factors of various systems have accordingly been revealed in order to facilitate process intensification, operational control, and even onsite monitoring. The major contributions of causality explorations are primarily in SPC and should be extended to other areas, such as material development, through the implementation of powerful *post-hoc* model-agnostic analyses.

One of the most challenging aspects to investigate is the ability to explain the inner rationales of AI systems. Nonetheless, AI explainability remains an effective way to enhance the reliability of AI applications in CE. The examination of the inner mechanisms of AI applications is crucial for accelerating AI transformations in CE and gaining valuable guidance for the development of data-driven algorithms.

This review aims to improve chemical engineers' awareness of the significance of developing reliable AI transformative utilizations within CE. Consequentially, it is anticipated that chemical engineers will reach a consensus on reliable AI utilization as an innovative technique with the potential to address bottleneck challenges in our field.

Compliance with ethics guidelines

Yue Yuan, Donovan Chaffart, Tao Wu, and Jesse Zhu declare that they have no conflict of interest or financial conflicts to disclose.

References

[1] Drai D. Why AI-driven analytics is essential for data-driven decision-making [Internet]. New York City: Forbes; 2021 Dec 27 [cited 2022 Apr 10]. Available from: https://www.forbes.com/sites/forbestechcouncil/2021/12/27/why-ai-driven-analytics-is-essential-for-data-driven-decision-making/?sh=61a1efa373f4.

[2] IDC forecasts companies to spend almost $342 billion on AI solutions in 2021 [Internet]. Needham: Businesswire; 2021 Aug 4 [cited 2024 Mar 26]. Available from: https://www.businesswire.com/news/home/20210804005239/en/IDC-Forecasts-Companies-to-Spend-Almost-342-Billion-on-AI-Solutions-in-2021.

[3] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 2019;9(4):e1312.

[4] 2024 AI business predictions [Internet]. London: PwC; [cited 2022 Apr 10]. Available from: https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html.

[5] Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: a review of machine learning interpretability methods. Entropy 2021;23(1):18.

[6] Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020;27(3):491–7.

[7] Miró-Nicolau M, Moyà-Alcover G, Jaume-i-Capó A. Evaluating explainable artificial intelligence for X-ray image analysis. Appl Sci 2022;12(9):4459.

[8] Sivaram A, Venkatasubramanian V. XAI-MEG: combining symbolic AI and machine learning to generate first-principles models and causal explanations. AIChE J 2022;68(6):e17687.

[9] Bhakte A, Pakkiriswamy V, Srinivasan R. An explainable artificial intelligence based approach for interpretation of fault classification results from deep neural networks. Chem Eng Sci 2022;250:117373.

[10] Schweidtmann AM, Esche E, Fischer A, Kloft M, Repke JU, Sager S, et al. Machine learning in chemical engineering: a perspective. Chemie Ingenieur Technik 2021;93(12):2029–39.

[11] Chakraborty A, Sivaram A, Venkatasubramanian V. AI-DARWIN: a first principles-based model discovery engine using machine learning. Comput Chem Eng 2021;154:107470.

[12] Shin D. The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. Int J Hum Comput Stud 2021;146:102551.

[13] Pirdashti M, Curteanu S, Kamangar MH, Hassim MH, Khatami MA. Artificial neural networks: applications in chemical engineering. Rev Chem Eng 2013;29(4):205–39.

[14] Zendehboudi S, Rezaei N, Lohi A. Applications of hybrid models in chemical, petroleum, and energy systems: a systematic review. Appl Energy 2018;228:2539–66.

[15] Venkatasubramanian V. The promise of artificial intelligence in chemical engineering: is it here, finally? AIChE J 2019;65(2):466–78.

[16] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.

[17] Wu H, Zhao J. Deep convolutional neural network model based chemical process fault diagnosis. Comput Chem Eng 2018;115:185–97.

[18] Turing AM. Computing machinery and intelligence. In: Epstein R, Roberts G, Beber G, editors. Parsing the Turing test. Dordrecht: Springer; 2009. p. 23–65.

[19] Vaidyanathan R, Venkatasubramanian V. On the nature of fault space classification structure developed by neural networks. Eng Appl Artif Intell 1992;5(4):289–97.

[20] Swartout WR, Moore JD. Explanation in second generation expert systems. In: David JM, Krivine JP, Simmons R, editors. Second generation expert systems. Berlin: Springer; 1993. p. 543–85.

[21] Minh D, Wang HX, Li YF, Nguyen TN. Explainable artificial intelligence: a comprehensive review. Artif Intell Rev 2022;55(5):3503–68.

[22] Gunning D, Vorm E, Wang JY, Turek M. DARPA's explainable AI (XAI) program: a retrospective. Appl AI Lett 2021;2(4):e61.

[23] Agarwal P, Tamer M, Budman H. Explainability: relevance based dynamic deep learning algorithm for fault detection and diagnosis in chemical processes. Comput Chem Eng 2021;154:107467.

[24] Panigutti C, Perotti A, Pedreschi D. Doctor XAI: an ontology-based approach to black-box sequential data classification explanations. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency; 2020 Jan 27–30; Barcelona, Spain. New York City: Association for Computing Machinery; 2020. p. 629–39.

[25] Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. 2017. arXiv:1702.08608.

[26] Miller T. Explanation in artificial intelligence: insights from the social sciences. Artif Intell 2019;267:1–38.

[27] Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020;58:82–115.

[28] Venkatasubramanian V. Teaching artificial intelligence to chemical engineers: experience from a 35-year-old course. Chem Eng Educ 2022;56(4):231–40.

[29] Maruyama Y. Symbolic and statistical theories of cognition: towards integrated artificial intelligence. In: Cleophas L, Massink M, editors. Software engineering and formal methods. Cham: Springer; 2021. p. 129–46.

[30] Ji W, Qiu W, Shi Z, Pan S, Deng S. Stiff-PINN: physics-informed neural network for stiff chemical kinetics. J Phys Chem A 2021;125(36):8098–106.

[31] Yang Y, Perdikaris P. Adversarial uncertainty quantification in physics-informed neural networks. J Comput Phys 2019;394:136–52.

[32] Yuan Y, Dong X, Ricardez-Sandoval L. Insights into syngas combustion on a defective NiO surface for chemical looping combustion: oxygen migration and vacancy effects. J Phys Chem C 2020;124(52):28359–70.

[33] Jung J, Gamwo IK. Multiphase CFD-based models for chemical looping combustion process: fuel reactor modeling. Powder Technol 2008;183(3):401–9.

[34] Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One 2015;10(7):e0130140.

[35] Konig R, Johansson U, Niklasson L. G-REX: a versatile framework for evolutionary data mining. In: Bonchi F, Berendt B, Giannotti F, Gunopulos D, Turini F, Zaniolo C, et al., editors. Proceedings of 2008 IEEE International Conference on Data Mining Workshops; 2008 Dec 15–19; Pisa, Italy. Piscataway: IEEE; 2008. p. 971–4.

[36] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: von Luxburg U, Guyon I, Bengio S, Wallach H, Fergus R, editors. Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017 Dec 4–9; Long Beach, CA, USA. Red Hook: Curran Associates Inc.; 2017. p. 4768–77.

[37] Hale WT, Safikou E, Bollas GM. Inference of faults through symbolic regression of system data. Comput Chem Eng 2022;157:107619.

[38] Agarwal P, Tamer M, Budman H. Assessing observability using supervised autoencoders with application to Tennessee Eastman process. IFAC-PapersOnLine 2020;53(2):206–11.

[39] Agarwal P, Tamer M, Sahraei MH, Budman H. Deep learning for classification of profit-based operating regions in industrial processes. Ind Eng Chem Res 2020;59(6):2378–95.

[40] Larsson T, Hestetun K, Hovland E, Skogestad S. Self-optimizing control of a large-scale plant: the Tennessee Eastman process. Ind Eng Chem Res 2001;40(22):4889–901.

[41] Downs JJ, Vogel EF. A plant-wide industrial process control problem. Comput Chem Eng 1993;17(3):245–55.

[42] Duan C, Liu F, Nandy A, Kulik HJ. Putting density functional theory to the test in machine-learning-accelerated materials discovery. J Phys Chem Lett 2021;12(19):4628–37.

[43] He J, Wang L, Zhang C, Sun W, Yin Z, Zhang H, et al. A high throughput screening model of solidophilic flotation reagents for chalcopyrite based on quantum chemistry calculations and machine learning. Miner Eng 2022;177:107375.

[44] Kim J, Gu GH, Noh J, Kim S, Gim S, Choi J, et al. Predicting potentially hazardous chemical reactions using an explainable neural network. Chem Sci 2021;12(33):11028–37.

[45] Kikutsuji T, Mori Y, Okazaki K, Mori T, Kim K, Matubayasi N. Explaining reaction coordinates of alanine dipeptide isomerization obtained from deep neural networks using explainable artificial intelligence (XAI). J Chem Phys 2022;156(15):154108.

[46] Langlotz CP, Allen B, Erickson BJ, Kalpathy-Cramer J, Bigelow K, Cook TS, et al. A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/The Academy Workshop. Radiology 2019;291(3):781–91.

[47] Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat Mach Intell 2020;2(10):573–84.

[48] Sallam IE, Abdelwareth A, Attia H, Aziz RK, Homsi MN, von Bergen M, et al. Effect of gut microbiota biotransformation on dietary tannins and human health implications. Microorganisms 2021;9(5):965.

[49] Ward IR, Wang L, Lu J, Bennamoun M, Dwivedi G, Sanfilippo FM. Explainable artificial intelligence for pharmacovigilance: what features are important when predicting adverse outcomes? Comput Methods Programs Biomed 2021;212:106415.

[50] Yang Z, Zhong W, Zhao L, Chen CYC. MGraphDTA: deep multiscale graph neural network for explainable drug–target binding affinity prediction. Chem Sci 2022;13(3):816–33.

[51] Das L, Sivaram A, Venkatasubramanian V. Hidden representations in deep neural networks: part 2. Regression problems. Comput Chem Eng 2020;139:106895.

[52] Sivaram A, Das L, Venkatasubramanian V. Hidden representations in deep neural networks: part 1. Classification problems. Comput Chem Eng 2020;134:106669.

[53] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A comprehensive survey on transfer learning. Proc IEEE 2021;109(1):43–76.

[54] Lu J, Behbood V, Hao P, Zuo H, Xue S, Zhang G. Transfer learning using computational intelligence: a survey. Knowl Base Syst 2015;80:14–23.

[55] Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. Nat Rev Phys 2021;3(6):422–40.

[56] Mao Z, Jagtap AD, Karniadakis GE. Physics-informed neural networks for high-speed flows. Comput Methods Appl Mech Eng 2020;360:112789.

[57] Wu JL, Xiao H, Paterson E. Physics-informed machine learning approach for augmenting turbulence models: a comprehensive framework. Phys Rev Fluids 2018;3(7):074602.

[58] Gao H, Sun L, Wang JX. PhyGeoNet: physics-informed geometry-adaptive convolutional neural networks for solving parameterized steady-state PDEs on irregular domain. J Comput Phys 2021;428:110079.

[59] Kashinath K, Mustafa M, Albert A, Wu JL, Jiang C, Esmaeilzadeh S, et al. Physics-informed machine learning: case studies for weather and climate modelling. Phil Trans R Soc A 2021;379(2194):20200093.

[60] Maheshwari S, Shetty S, Ratnakar R, Sanyal S. Role of computational science in materials and systems design for sustainable energy applications: an industry perspective. J Indian Inst Sci 2022;102(1):11–31.

[61] Kashefi A, Rempe D, Guibas LJ. A point-cloud deep learning framework for prediction of fluid flow fields on irregular geometries. Phys Fluids 2021;33(2):027104.

[62] Lu L, Jin P, Pang G, Zhang Z, Karniadakis GE. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. Nat Mach Intell 2021;3(3):218–29.

[63] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. In: Arbib MA, editor. The handbook of brain theory and neural networks. Cambridge: MIT Press; 1995. p. 255–8.

[64] Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, et al. Graph neural networks: a review of methods and applications. AI Open 2020;1:57–81.

[65] Raissi M, Perdikaris P, Karniadakis GE. Numerical Gaussian processes for time-dependent and nonlinear partial differential equations. SIAM J Sci Comput 2018;40(1):A172–98.

[66] Erichson NB, Muehlebach M, Mahoney MW. Physics-informed autoencoders for Lyapunov-stable fluid flow prediction. 2019. arXiv:1905.10866.

[67] Geneva N, Zabaras N. Transformers for modeling physical systems. Neural Netw 2022;146:272–89.

[68] Pang G, Lu L, Karniadakis GE. fPINNs: fractional physics-informed neural networks. SIAM J Sci Comput 2019;41(4):A2603–26.

[69] Yang L, Zhang D, Karniadakis GE. Physics-informed generative adversarial networks for stochastic differential equations. SIAM J Sci Comput 2020;42(1):A292–317.

[70] Lu L, Meng X, Mao Z, Karniadakis GE. DeepXDE: a deep learning library for solving differential equations. SIAM Rev 2021;63(1):208–28.

[71] Wang Y, Han X, Guo D, Lu L, Chen Y, Ouyang M. Physics-informed recurrent neural networks with fractional-order constraints for the state estimation of lithium-ion batteries. Batteries 2022;8(10):148.

[72] Yuan L, Ni YQ, Deng XY, Hao S. A-PINN: auxiliary physics informed neural networks for forward and inverse problems of nonlinear integro-differential equations. J Comput Phys 2022;462:111260.

[73] Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys 2019;378:686–707.

[74] Raissi M, Perdikaris P, Karniadakis GE. Physics informed deep learning (part I): data-driven solutions of nonlinear partial differential equations. 2017. arXiv:1711.10561.

[75] Raissi M, Perdikaris P, Karniadakis GE. Physics informed deep learning (part II): data-driven discovery of nonlinear partial differential equations. 2017. arXiv:1711.10566.

[76] Cai S, Mao Z, Wang Z, Yin M, Karniadakis GE. Physics-informed neural networks (PINNs) for fluid mechanics: a review. Acta Mech Sin 2021;37:1727–38.

[77] Cai S, Li H, Zheng F, Kong F, Dao M, Karniadakis GE, et al. Artificial intelligence velocimetry and microaneurysm-on-a-chip for three-dimensional analysis of blood flow in physiology and disease. Proc Natl Acad Sci USA 2021;118(13):e2100697118.

[78] Quintino AM, da Rocha DLLN, Fonseca Jr R, Rodriguez OMH. Flow pattern transition in pipes using data-driven and physics-informed machine learning. J Fluids Eng 2021;143(3):031401.

[79] De Florio M, Schiassi E, Furfaro R. Physics-informed neural networks and functional interpolation for stiff chemical kinetics. Chaos 2022;32(6):063107.

[80] Ngo SI, Lim YI. Solution and parameter identification of a fixed-bed reactor model for catalytic CO_2 methanation using physics-informed neural networks. Catalysts 2021;11(11):1304.

[81] Alhajeri MS, Abdullah F, Wu Z, Christofides PD. Physics-informed machine learning modeling for predictive control using noisy data. Chem Eng Res Des 2022;186:34–49.

[82] Zeng X, Xue CD, Chen KJ, Li YJ, Qin KR. Deep-learning-assisted extraction of height-averaged velocity from scalar signal transport in a shallow microfluidic channel. Microfluid Nanofluidics 2022;26(5):36.

[83] Afzal MAF, Sonpal A, Haghighatlari M, Schultz AJ, Hachmann J. A deep neural network model for packing density predictions and its application in the study of 1.5 million organic molecules. Chem Sci 2019;10(36):8374–83.

[84] He Z, Ni F, Wang W, Zhang J. A physics-informed deep learning method for solving direct and inverse heat conduction problems of materials. Mater Today Commun 2021;28:102719.

[85] Shih YT, Shi Y, Huang L. Predicting glass properties by using physics- and chemistry-informed machine learning models. J Non-Cryst Solids 2022;584:121511.

[86] Song W, Du W, Fan C, Yang M, Qian F. Adaptive weighted hybrid modeling of hydrocracking process and its operational optimization. Ind Eng Chem Res 2021;60(9):3617–32.

[87] Rajulapati L, Chinta S, Shyamala B, Rengaswamy R. Integration of machine learning and first principles models. AIChE J 2022;68(6):e17715.

[88] Arrieta-Escobar JA, Bernardo FP, Orjuela A, Camargo M, Morel L. Incorporation of heuristic knowledge in the optimal design of formulated products: application to a cosmetic emulsion. Comput Chem Eng 2019;122:265–74.

[89] Schubert J, Simutis R, Dors M, Havlík I, Lübbert A. Hybrid modelling of yeast production processes—combination of *a priori* knowledge on different levels of sophistication. Chem Eng Technol 1994;17(1):10–20.

[90] Kosko B. Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence. Upper Saddle River: Prentice-Hall, Inc.; 1992.

[91] Yuan Y, Dong X, Ricardez-Sandoval L. A multi-scale model for syngas combustion on NiO oxygen carrier for chemical looping combustion: the role of nearest neighbors. Fuel Process Technol 2022;229:107172.

[92] Yuan Y, Dong X, Ricardez-Sandoval L. A multi-scale simulation of syngas combustion reactions by Ni-based oxygen carriers for chemical looping combustion. Appl Surf Sci 2020;531:147277.

[93] Sansana J, Joswiak MN, Castillo I, Wang Z, Rendall R, Chiang LH, et al. Recent trends on hybrid modeling for Industry 4.0. Comput Chem Eng 2021;151:107365.

[94] Petch J, Di S, Nelson W. Opening the black box: the promise and limitations of explainable machine learning in cardiology. Can J Cardiol 2022;38(2):204–13.

[95] Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, et al. A survey on deep learning: algorithms, techniques, and applications. ACM Comput Surv 2019;51(5):92.

[96] Von Stosch M, Oliveira R, Peres J, Feyo de Azevedo S. Hybrid semi-parametric modeling in process systems engineering: past, present and future. Comput Chem Eng 2014;60:86–101.

[97] Von Stosch M, Davy S, Francois K, Galvanauskas V, Hamelink JM, Luebbert A, et al. Hybrid modeling for quality by design and PAT—benefits and challenges of applications in biopharmaceutical industry. Biotechnol J 2014;9(6):719–26.

[98] Ulissi ZW, Medford AJ, Bligaard T, Nørskov JK. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. Nat Commun 2017;8(1):14621.

[99] Peterson AA. Acceleration of saddle-point searches with machine learning. J Chem Phys 2016;145(7):074106.

[100] Rochac JFR, Zhang N, Thompson L, Oladunni T. A data augmentation-assisted deep learning model for high dimensional and highly imbalanced hyperspectral imaging data. In: Proceedings of 2019 9th International Conference on Information Science and Technology; 2019 Aug 2–5; Hulunbuir, China. Piscataway: IEEE; 2019. p. 362–7.

[101] Ponce H, de Campos Souza PV, Guimarães AJ, Gonzalez-Mora G. Stochastic parallel extreme artificial hydrocarbon networks: an implementation for fast and robust supervised machine learning in high-dimensional data. Eng Appl Artif Intell 2020;89:103427.

[102] Sharma N, Liu Y. A hybrid science-guided machine learning approach for modeling chemical processes: a review. AlChE J 2022;68(5):e17609.

[103] Fang H, Zhou J, Wang Z, Qiu Z, Sun Y, Lin Y, et al. Hybrid method integrating machine learning and particle swarm optimization for smart chemical process operations. Front Chem Sci Eng 2022;16(2):274–87.

[104] Khoei AR, Keshavarz S, Biabanaki SOR. Optimal design of powder compaction processes via genetic algorithm technique. Finite Elem Anal Des 2010;46 (10):843–61.

[105] González-García R, Rico-Martínez R, Kevrekidis IG. Identification of distributed parameter systems: a neural net based approach. Comput Chem Eng 1998;22(Suppl 1):S965–8.

[106] Yang Z, Lu B, Wang W. Coupling artificial neural network with EMMS drag for simulation of dense fluidized beds. Chem Eng Sci 2021;246:117003.

[107] Pan I, Mason LR, Matar OK. Data-centric engineering: integrating simulation, machine learning and statistics. Challenges and opportunities. Chem Eng Sci 2022;249:117271.

[108] Psichogios DC, Ungar LH. A hybrid neural network—first principles approach to process modeling. AlChE J 1992;38(10):1499–511.

[109] Nielsen RF, Nazemzadeh N, Sillesen LW, Andersson MP, Gernaey KV, Mansouri SS. Hybrid machine learning assisted modelling framework for particle processes. Comput Chem Eng 2020;140:106916.

[110] Schenker B, Agarwal M. Online-optimized feed switching in semi-batch reactors using semi-empirical dynamic models. Control Eng Pract 2000;8 (12):1393–403.

[111] Guan K, Gao Y, Zeng Q, Luan X, Zhang Y, Cheng L, et al. Numerical modeling of SiC by low-pressure chemical vapor deposition from methyltrichlorosilane. Chin J Chem Eng 2020;28(6):1733–43.

[112] Lu X, Deng J, Xiao Y, Zhai X, Wang C, Yi X. Recent progress and perspective on thermal-kinetic, heat and mass transportation of coal spontaneous combustion hazard. Fuel 2022;308:121234.

[113] Niu Z, Pinfield VJ, Wu B, Wang H, Jiao K, Leung DYC, et al. Towards the digitalisation of porous energy materials: evolution of digital approaches for microstructural design. Energy Environ Sci 2021;14(5): 2549–76.

[114] Nikolopoulos A, Samlis C, Zeneli M, Nikolopoulos N, Karellas S, Grammelis P. Introducing an artificial neural network energy minimization multi-scale drag scheme for fluidized particles. Chem Eng Sci 2021;229:116013.

[115] Chaffart D, Ricardez-Sandoval LA. Optimization and control of a thin film growth process: a hybrid first principles/artificial neural network based multiscale modelling approach. Comput Chem Eng 2018;119:465–79.

[116] Vvedensky DD, Zangwill A, Luse CN, Wilby MR. Stochastic equations of motion for epitaxial growth. Phys Rev E 1993;48(2):852–62.

[117] Yuan Y, Zhu J. Intelligent intercommunicating multiscale engineering: the engineering of the future. Engineering 2023;30:13–9.

[118] You H, Yuan Y, Li J, Sandoval LR. A multi-scale model for CO_2 capture: a nickel-based oxygen carrier in chemical-looping combustion. IFAC-PapersOnLine 2018;51(18):97–102.

[119] de Campos Souza PV. Fuzzy neural networks and neuro-fuzzy networks: a review the main techniques and applications used in the literature. Appl Soft Comput 2020;92:106275.

[120] Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D 2020;404:132306.

[121] Wen L, Li X, Gao L, Zhang Y. A new convolutional neural network-based data-driven fault diagnosis method. IEEE Trans Ind Electron 2018;65(7): 5990–8.

[122] Perera ATD, Kamalaruban P. Applications of reinforcement learning in energy systems. Renew Sustain Energy Rev 2021;137:110618.