TU Delft

Delft University of Technology

Document Version
Final published version

Licence
Dutch Copyright Act (Article 25fa)

Citation (APA)
Kiss, A. A., & Grievink, J. (2026). Perspectives on the Essential Role of First-Principles Modeling in the Age of AI. *Industrial and Engineering Chemistry Research*, *65*(8), 4234-4253. https://doi.org/10.1021/acs.iecr.5c04156

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

I&EC research
Industrial & Engineering Chemistry Research
pubs.acs.org/IECR
Perspective

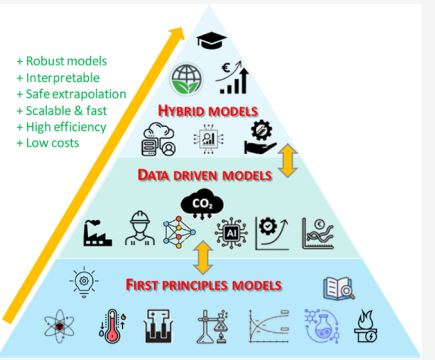
Perspectives on the Essential Role of First-Principles Modeling in the Age of AI

Published as part of Industrial & Engineering Chemistry Research special issue "Importance of First-Principles Modeling in the Era of Growing AI".

Anton A. Kiss* and Johan Grievink

Cite This: *Ind. Eng. Chem. Res.* 2026, 65, 4234–4253
Read Online
ACCESS | Metrics & More | Article Recommendations

ABSTRACT: The evolution of artificial intelligence (AI), machine learning (ML), and neural networks (NN) is transforming the landscape of scientific and engineering modeling. It also prompts a debate on the role of first-principles modeling (FPM) in chemical engineering. While data-driven methods excel at interpolation and very rapid development, they often lack physical fidelity, interpretability, and reliable extrapolation capabilities. This article provides a personal academic and industrial perspective on the synergistic integration of FPM and AI-based methods, highlighting their complementary roles in process systems engineering. We argue that FPM (based on fundamental conservation laws and mechanistic understanding of phenomena), remains indispensable for ensuring robustness, safety, physical consistency, and adaptability of models in PSE. Moreover, we analyze the synergistic potential of hybrid approaches by deconstructing the model-building workflow. The latter is the primary lens to identify key decision points where integration delivers maximum value, moving beyond a simple paradigm comparison. Using this structured analysis of the model-building workflow, we identify several major opportunities for this integration, particularly where first-principles knowledge is incomplete. The discussion extends to practical strategies for model validation, scalability, and industrial applications, supported by case studies, as well as the potential of LLMs in assisting the future developments of FPM. Finally, we conclude that a physics-informed foundation for modeling is not obsolete but is instead critical for guiding the safe and reliable application of AI in chemical engineering.



Downloaded via TU DELFT on March 26, 2026 at 13:49:04 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

1. INTRODUCTION

1.1. Purpose and Scope

Models are at the core of process systems engineering (PSE) tasks such as process design, control, optimization, intensification, scheduling, and supply chain management.[1−3] However, the building of a model serves one or more purposes, as a model is not a goal in itself. A model relates to a chemical engineering object, either existing in physical reality or at a conceptual stage of development. A model mimics aspects of behavior, structure and performance of an object in a digital space. The purpose shapes a model and its degree of complexity or simplicity. Typically, these models have beyond their specific application purposes also the role of transferable knowledge carriers, particularly if the physical principles underlying the model are kept transparent in the structure of the equations.[2] Setting up a model provides a structured approach to understanding the system described, often across many scales in time and space and over a substantial part of the life span of its object. First-principles models (FPM) based on descriptions of physical and (bio-) chemical phenomena that change the states of physical resources, have been very successful over the past half century.[4−9]

However, Artificial Intelligence (AI) is transforming the landscape of science and engineering, with many recent advances pointing as strong evidence of its impact.[10] This shift toward a data-centric approach has led to increased reliance on empirical and so-called black-box models in chemical engineering—see Figure 1 for context. In the view of the authors (with both academic research and industrial experience), a realistic trajectory for the coming decade involves growth in data-driven methods' application domains, rather than the replacement of physics-based approaches. This perspective is based on the recognition that different application contexts favor different approaches. FPM continues to be essential, providing founda-

Received: October 5, 2025
Revised: January 24, 2026
Accepted: February 11, 2026
Published: February 19, 2026

ACS Publications
© 2026 American Chemical Society

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

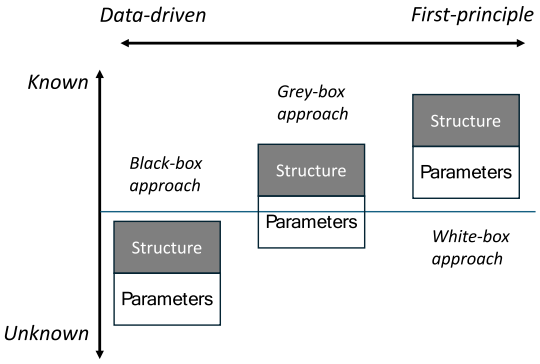


Figure 1. Data-driven vs first principle models: a shift in known/unknown structure and parameters.

tional understanding, maintaining physical consistency, and allowing extrapolation far beyond the bounds of available data. The real question is *to what extent will data-driven modeling expand in chemical engineering application domains, and how should organizations strategically integrate physics-based and data-driven approaches to optimize value creation, innovation, and operational performance*?

While academic research enjoys quite significant funding to explore DDM's possibilities, the industrial adoption is more measured and pragmatic. Industry seems to be actively pursuing modeling innovation, driven by economic pressures and data availability. However, the adoption patterns are sector-dependent and application-specific. Pharmaceutical, biotech, and safety-critical sectors prioritize FPM due to regulatory requirements. However, mature commodities and downstream operations increasingly use DDM for predictive maintenance, soft-sensing, and real-time optimization. As computational power increases and fundamental models become more manageable at scale, we anticipate that increasingly sophisticated hybrid approaches will emerge. Physics-based models may transition from serving as standalone tools to becoming components within integrated frameworks that leverage both first-principles understanding and data-driven acceleration. This natural evolution reflects complementarity rather than replacement.

Invited for the special issue of *Industrial & Engineering Chemistry Research* about *Importance of first-principles modeling in the era of growing AI* − this perspective paper offers a lightly referenced scholarly opinion about the symbiotic relationship between FPM and AI, emphasizing that their judicious integration offers a powerful path forward. The focus is on the essential and continuing role of physics-based modeling (in process design, scale-up, industrial innovation, and the development of robust process understanding), while fully acknowledging the mature and widespread success of data-driven methods in confined windows for operations and control. We discuss strategies for leveraging the strengths of each approach, address limitations, and highlight the importance of integrating physics-based insights into modern computational frameworks. This paper moves beyond a simple comparison to analyze the model development lifecycle. We use this framework to identify the specific strengths and limitations of each paradigm and the points where their integration creates maximum value. This leads to a more nuanced understanding than the prevailing FPM vs AI portrayal, advocating actually for an FPM and AI synergy.

1.2. On the Rise of AI and DDM in Chemical Engineering

AI has a broad historical emerging presence, with several successive AI waves: 1) Expert systems and early ML (1980s-2000s); 2) Big Data and classical deep learning (2010s), and 3) Generative AI era (2022-present) with LLMs and foundation models. In this work "age of AI" specifically refers to the most recent wave (not AI broadly), namely a recent surge in Generative AI capabilities (e.g., LLMs such as ChatGPT, Gemini, Claude, DeepSeek; diffusion models; large multimodal models) that have sparked renewed debates about FPM's role and relevance. Machine learning (ML) algorithms, neural networks, and large-scale data processing approaches remain at the core of many current developments, allowing models to uncover rather complex patterns within large data sets.[11] This study addresses two distinct topics: 1) *Classical DDM/ML* (neural networks, support vector machines, ensemble methods, and other function approximators for input-output modeling of process variables) as this represents ∼80% of current industrial DDM, and 2) *Generative AI/LLM* (foundation models with emergent capabilities for text generation, reasoning, and scientific hypothesis formation) as this represents an emerging frontier for PSE. While both fall under the broad AI umbrella, they present different characteristics regarding explainability, trustworthiness, physical consistency, and educational implications for chemical engineers. There are also clear distinctions between classical data-driven models and Generative AI. In case of classical DDM, the function approximators are trained on specific data sets to predict system behavior within the training domain. Thus, their operation is deterministic given fixed sets of inputs; transparency issues reflect the high-dimensional nature of learned representations, not inherent nondeterminism. For Generative AI/LLMs, the foundation models are trained on massive data, exhibiting emergent reasoning capabilities including language generation, analogical reasoning, and hypothesis formation. Their probabilistic nature and broad training basis create qualitatively different challenges and opportunities for scientific application. Therefore, the key distinction is that while classical DDM requires extensive experimental data for a specific system, LLMs bring pretrained scientific knowledge and reasoning patterns that could theoretically assist in FPM generation. However, this potential is coupled with unprecedented challenges in ensuring physical validity.

Nowadays there are well-established DDM niches in chemical engineering, such as 1) *Model Predictive Control (MPC)* using data-driven models which dominate real-time control in large refineries and petrochemical plants; 2) *Partial Least Squares (PLS)* as the industry standard for soft-sensing applications with proven reliability; and 3) *Principal Component Analysis (PCA)* which is the de facto monitoring standard for plant-wide anomaly detection. They are pillars of mature industrial practice.[10,12,13] However, DDM adoption remains more uncertain for innovation-intensive design tasks requiring safe extrapolation (e.g., novel reactor synthesis, scale-up under untested conditions), where regulatory and safety constraints favor FPM. DDM has established well-developed application domains where it delivers significant industrial value, particularly in real-time operations, process monitoring, and soft-sensing. These matured application domains are standard industrial practices in modern chemical plants.

Traditional FPM approaches remain the golden standard across industry, often forming the solid base upon which DDM enhancements are layered.[14] Recent authoritative sources[15−17] confirm FPM's continued role in modern process design. This gap in adoption highlights the different priorities at play: while academia is frequently drawn to novel computational achieve-

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

ments, industry is more focused on robustness, safety, reliability, interpretability, and the total cost of ownership across the model's lifecycle. This is not just an academic debate as the decision between methods is often driven by the full spectrum of costs: development, validation, computational expense, human interpretability, and ongoing maintenance. From this broader life-cycle perspective, while DDM may seem less expensive to develop initially, long-term costs (around validation, adaptation, and interpretation) can rise quickly, particularly when process conditions shift or structures evolve. The recent surge in DDM is enabled by an increased availability of data from long-term digital storage in plants and laboratories, as well as a proliferation of new sensor and measurement technologies. However, note that the data used for training academic DDM studies is often not measured plant data but rather simulation data generated from existing, validated FPMs (synthetic "low fidelity" data). While useful for method development, this practice can create a misleading perception of data abundance and overlooks the significant challenges of obtaining high-fidelity, informative experimental data for real-world validation.

Despite all the successes of DDM, FPM remains irreplaceable alongside Generative AI for several key reasons:

Physical guardrails, as Generative AI inherently lacks mechanisms to enforce conservation laws (mass, energy, momentum) or thermodynamic consistency, which are crucial for safety-critical applications. For this reason, many failure scenarios (such as predicting negative mass fractions, violating enthalpy balances, using perpetual-motion energy cycles) could cause catastrophic failures in real plants.

Causality vs correlation, since FPM provides mechanistic causal explanations enabling root-cause analysis (e.g., "why did fouling occur?"), while Generative AI identifies sensitivity patterns and correlations without mechanistic insight. In regulated industries (such as pharma and aerospace) and innovation scenarios, mechanistic understanding is non-negotiable.

Data efficiency and extrapolation, as FPM requires minimal data for extrapolation beyond training conditions (n.b. extrapolation is included into physics), while Generative AI typically requires massive data sets and struggles with extrapolation. FPM provides reliable predictions for new processes, extreme conditions, or rare events, whereas AI fails catastrophically.

Regulatory compliance and transparency, since FPM satisfies stringent industry regulations (FDA process validation, HSE risk assessment) where black-box AI models face regulatory skepticism. The ability to justify predictions through physics provides the audit trail regulators demand.

2. FIRST-PRINCIPLES MODELING IN THE CONTINUUM DOMAIN

First-principles models (FPM), also known as white-box/transparent-box models (Figure 1), are mathematical representations of systems based on fundamental physical laws and principles. They aim to capture the underlying mechanisms and relationships within a system, rather than simply learning patterns from data (as black-box or data-driven models do). The approach used in FPM is valuable in situations where understanding the physical behavior of a system is crucial, such as in chemical engineering applications related to product and process design, process scale-up and process control and optimization. A chemical engineering model is based on physical principles if such a model contains:

- *Distinctive spatial domains and time regimes*. Matter is present in any domain, while matter is characterized by the identities of chemical species, physical phase(s) and physical state. The state of matter can change due to physical events and chemical reactions.
- *Variables per spatial domain* that relate to the physical states of matter (gas, liquid, solids), physical resources (mass, energy, momentum, electrical charge), thermodynamic state variables (such as density, temperature, pressure, concentrations of species), physical flow fields and fluxes for the transport of physical resources and rates of chemical reactions.
- *Equations per spatial domain* for conservation of physical resources (such as mass, energy, momentum, and electrical charge); the interactions of electrical, magnetic or gravitational fields with matter; physical state of matter in distinctive phases; physical transport phenomena and/or chemical reaction rates; and changes in the number and properties of particles having distributive properties.

To be fair, physics-based models (as used in chemical engineering) inherently contain both deductive elements (such as conservation laws, thermodynamics) and inductive elements (such as empirical rate laws, property correlations), making them fundamentally hybrid in nature. This means that no engineering model is purely first-principles, as all models contain approximations and empirical elements.

Moreover, modeling approaches are lying on a spectrum rather than in discrete categories, going from purely physics-based (conservation laws only), through physics-based with empirical correlations, then hybrid models with physics-informed components, data-driven models with soft constraints, to black-box data-driven models. Rather than treating hybrid models as a completely separate category, we view them as strategic choices about where on the physics-to-data-driven continuum to position the model for a specific application purpose. Having clarified these distinctions, this section focuses on the importance of strategies for effective use of FPMs.

2.1. Fundamental Role and Physical Correctness

Many chemical engineering systems extend over a wide range of scales in time and space. Commonly, three main clusters of scales are considered in chemical engineering modeling:

1 *Tiny scales*: Statistical modeling of big populations of molecules or other particle-like entities (domain of *Quantum Chemistry* and *Statistical Mechanical* modeling).
2 *Medium scales*: Continuum domain, including physical transport phenomena, reactions, continuously distributed particle populations, multiphase products, and process unit operations.
3 *Large scales*: Networks with multiple nodes that are interconnected for processing resources (process flowsheets, process plants, connected by supply chains).

While acknowledging the relevance of these three clusters for the chemical engineering profession, the focus of this article is on models in the continuum domain.

First-principles based models, more accurately termed physics-based or mechanistic models, are mathematical representations grounded in fundamental conservation laws, thermodynamic principles, and empirically validated constitutive relations describing transport phenomena and reaction rates. While the conservation laws and thermodynamic framework represent first-principles foundations, the constitutive

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

equations themselves often include semiempirical correlations and parameter values fitted to experimental data, thus including a data-informed component within an overall physics-based structure. Together all the equations ensure that models adhere to physical reality, although some (realistic) assumptions are used for simplification of modeling especially for momentum equations (e.g., plug flow in pipes or well mixed tanks). This adherence is essential for safety-critical applications, process scale-up, operation, and optimization tasks where empirical models may fail. For example, thermodynamic consistency guarantees that phase equilibrium calculations respect the fundamental laws, thus preventing erroneous unphysical predictions. Conservation laws are hard constraints that enhance model reliability and improve interpretability.[14]

However, many first-principles models also have a (somewhat hidden) data-driven edge. That is related to the quantitative descriptions of transport phenomena, phase equilibria and reaction rates. Most chemical engineering models contain so-called constitutive equations (e.g., thermodynamic equations of state; rate laws for reactions and for transfers of mass, heat, momentum, and electrical charge). These constitutive equations often have a semiempirical nature, implying that the structure of such equations is often compatible with thermodynamic insights about relevant driving forces. Yet, there are parameters in these equations translating the magnitudes of the prevailing forces into rates of change of the amount and the state of a physical resource. One may extract numerical values for such parameters from more fundamental models at a deeper level (e.g., statistical or quantum mechanical models). Such avenues are now becoming available by multiscale modeling.[18,19] However, it is still common to obtain numerical parameter estimates from measured data (within a well-defined window of experimental conditions) by means of statistical parameter fitting methods. This practice of fitting model parameter to available experimental data introduces implicitly limited ranges of applicability of constitutive equations. These ranges are linked to the ranges over which experimental data for validation is available. In addition, such parameter fitting procedures introduce uncertainties in model predictions. These originate from statistical variabilities in the model parameter values propagating through the model equations into statistical uncertainties in computed model outputs.[20,21]

Transfer of chemical engineering knowledge in education and engineering practice relies heavily on the use of FPMs. This is the case in teaching key topics, such as process modeling,[8,22] process design and engineering,[9,23−27] process dynamics, control and optimization.[28−30] It is hard to imagine how to explain without the use of FPMs the impact of chemical and physical phenomena in process units and products on the technological and operational behavior and performance of a process plant.

2.2. Generic Features of Model Behavior

First-principles models of a chemical engineering system possess a generic structure, involving physical inputs, states and outputs. Furthermore, a model has parameters in its equations, some of which may have uncertain values. The physical state relates to amounts of species mass, energy and momentum. To bridge physics-based modeling and control theory, Kalman[31] introduced critical systemic properties for system analysis. These properties are also instrumental to the evaluation of first-principles models:

- *Observability*: Can the internal dynamic state be reconstructed from observations of system outputs?
- *Controllability*: Can the internal dynamic state be controlled by manipulating system inputs?
- *Identifiability*: Can unknown model parameters be uniquely determined from observed inputs and outputs?
- *Stability*: Will a dynamic state attain a stationary, finite value for constant inputs?

For process systems operating at a steady state, two additional properties are essential:

- *Conservation*: Physical entities (e.g., mass of elements, energy) must be conserved as they are in nature.
- *Multiplicity*: Nonlinear algebraic equations can lead to multiple solutions with different stability features, while for nonlinear dynamic systems bifurcations can arise.

First-principles models have physically meaningful internal states and are inherently amenable to analysis of these features. This contrasts with purely data-driven input-output models, for which concepts such as observability, controllability, and internal stability are often undefined, while ensuring conservation remains a major challenge. Identifiability can always be ascertained by assessing the rank of the information matrix of a parameter estimation problem $[X^T\Sigma^{-1}X]$, where X is the matrix of derivatives of the response functions with respect to the unknown parameters and Σ is the covariance matrix of the measurement errors in the experimental response variables.[32]

2.3. Defining Validity Domains

Understanding where a model applies is decisive in practice. Establishing validity domains is an essential step in the development and application of models, aimed to ensure that their predictions are reliable within specified conditions (e.g., range of temperature, pressure, concentrations, etc.). This activity involves identifying the range of input parameters, range of scales in time and space for which the model equations are valid, environmental conditions, and the assumptions under which the model accurately represents the physical system. In chemical engineering modeling the chosen form of the model balance equations is often in a space-time continuum, filled with mass, energy and momentum, as practiced in the domain of physical transport phenomena and chemical reaction engineering. However, at tinier scales one must often resort to a particle-based description, having populations of interacting particles, by statistical mechanics. Or even down to quantum mechanical descriptions. At a coarser scale, lumping is applied resulting in input-output descriptions for relevant resources per process unit operation, or per process in a supply chain network. Obviously, the resulting equations will be different.

Guarding model validity may even require adding a diagnostic set of inequality constraints to the variables that are present in model correlations for rate processes and physical properties. Inequality constraints on such variables represent also the limited ranges for which these correlations have been validated. Violation of such constraints in obtained solutions by model computations raises a red flag about a suspect solution. Hence, FPM enables explicit demarcation of domain validity based on physical principles, reducing risks associated with unphysical extrapolations. Model validation strategies (such as comparing predictions with experimental data, residual analysis, and sensitivity studies) are essential for establishing confidence. Therefore, defining validity domains is important for several reasons:[33−35]

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

- *Ensuring reliability*, as applying a model outside its validity domain can lead to unphysical or erroneous predictions, potentially causing costly mistakes or safety issues.
- *Guiding model use*, by clearly specifying where a model applies, thus aiding practitioners to select the appropriate tools for specific scenarios.
- *Facilitating model improvement*, as properly understanding its limitations informs targeted refinement and extension of models. In other words by *Pieter Schmal* (from ExxonMobil, and formerly from TU Delft) "*wrongness informs usefulness*" and this gives opportunities for improving models.

By explicitly defining the validity domain, modelers can avoid applying models to improper conditions where the model assumptions break down. This proactive approach minimizes the risk of wrong predictions, enhances confidence in model output, and supports responsible decision-making in chemical engineering applications.

2.4. Scalability and Computational Efficiency

High-fidelity models based on fundamental physical laws tend to be computationally intensive due to various reasons: e.g., fine spatial and temporal discretizations, complex nonlinearities arising from interactions between phenomena, counter-current contacting patterns between interacting thermodynamic phases, or large parameter spaces. This often limits their use to small-scale problems or requires significant computational resources. When trying to make FPM more scalable and efficient, it is all about striking the right balance between model fidelity and the computational resources. Techniques such as reduced-order modeling and surrogate models can help run complex simulations much faster (while keeping the physics intact by physics-informed approaches), thus enabling optimization and real-time decision-making. Some approaches to improve the scalability and efficiency of FPM include:

Reduced-order modeling which simplifies a high-fidelity model by projecting its behavior onto a lower-dimensional subspace that captures the dominant dynamics.[36] This is not just as a good technical solution but as an economic one as well.

Surrogate models (meta-models), which develop fast-to-evaluate approximations of the original complex model using machine learning, polynomial chaos, or regression techniques.[37]

Modular and hierarchical approaches that break down complex systems into interconnected modules or layers, each modeled at an appropriate fidelity level.

Physics-informed reduced models that include physical laws directly into the reduced models to ensure physical consistency and stability.

Parallel computing and high-performance computing that make use of distributed computing resources to accelerate simulations, especially for large-scale problems.

All these strategies transform prohibitive high-fidelity models into practical tools for large-scale and real-time applications without compromising their physical integrity.[36,37]

A common form of incremental scalability is in process intensification, leading to reductions in equipment size and/or increased functionality because of more active agents, improved contacting patterns between phases, and better operating conditions, all for the same processing function. Each one of such improvements can be related to specific parts in an FPM and be adapted. Also, new features can be easily added because there are clear relationships between model parts and physical features in the object of modeling.

It is worthwhile to properly analyze and trade-off the costs in different stages of the life span of a model. In engineering practice, one must account for the time and cost of obtaining computational solutions as well as for the time needed for assessments of obtained solutions by model users. A cumulative reduction in these repetitive costs may offset the one-time cost of human labor to reduce a model and to enhance its computational speed up-front. This life-cycle cost perspective is crucial when comparing FPM to DDM, as the latter may have lower initial development costs but higher long-term validation and adaptation costs, especially when process structure and/or conditions change.

2.5. Transparency, Interpretability, and Adaptability

First-principles models have inherent interpretability as models are based on physical concepts and equations. This transparency fosters trust, simplifies debugging, and helps with knowledge discovery, thus making FPM very valuable in scenarios where it is essential to understand the reasoning behind the model. Conversely, despite their predictive power, black-box ML models come with challenges related to transparency, validation, and acceptance.[38] That underlines the importance of balancing the predictive performance with interpretability, depending on the application context. In heavily regulated industries (e.g., healthcare, aerospace, or energy) the lack of transparency of black-box models can be a barrier to regulatory approval, as authorities require clear reasoning and validation of decision-making processes to ensure safety and compliance.[39−41]

FPM's transparency enables interpretability, but this potential is only realized under specific conditions:

Full transparency, as mathematical transparency alone is insufficient. Models must include equation-aware visualization tools, list of assumptions, sensitivity analysis, and explicit domain-of-validity demarcation.

Open preconfigured models, since commercial FPM packages often present black-box preconfigured models (property packages, kinetics correlations) whose internal structure is not transparent to the user. Therefore, users employ these without understanding their underpinnings.

Explicit domains of validity. FPM's transparency advantage is amplified when models include explicit inequality constraints on variables, and warn users when predictions venture beyond validated regions.

Documentation and training remain essential, as even FPM can be used blindly, without support. Thus, FPM's transparency depends on organizational commitment to knowledge transfer. FPM's structure enables transparency in ways black-box DDM cannot, even when actual transparency requires organizational effort.

Ideally, FPMs having empirical correlations should also be provided with associated domains of validity. These can be sets of (inequality) bounds within which the correlations are valid (experimentally tested). If computed answers in simulations show violations of some bounds, these violations should be reported. Possibly, such bound violations can be included in the visualization tools. Having explicit domains of validity is even more vital for optimization studies where models are pushed to extreme conditions against inequality constraints. That means that knowledge about local behavior reaches its limit.

The labor costs of developing a new model are substantial. Being able to extend its application life by making changes and extensions to a model in a transparent way is often less expensive than a new model development. Hence, adaptability of a model

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

in extending its domain of application is an attractive feature enhancing its practical lifespan. Whenever new relevant knowledge becomes available, a model can gradually be extended to more comprehensive versions. By virtue of the physical transparency of first-principles models, it is relatively easy to add or remove chemical species with their species (mass) balances. It is also transparent how to add or remove physical phenomena with local rate laws, or how to selectively adjust the various constitutive equations for the presence of new species. Such adaptability enhances the continued use of a model by its customers.

2.6. Limitations and Challenges

While FPM offers a robust framework for understanding complex systems (by relying on fundamental physical laws) it also comes with certain limitations and challenges. There are a few key issues that can hinder the effectiveness and applicability of FPM in real-world scenarios:

Complex systems with poorly understood mechanisms, as real-world systems involve intricate interactions and processes that are not fully understood or are difficult to model accurately. Generally, gaps in modeling can arise by 1) Wrongfully ignoring an active resource (like a reactive chemical species in trace amounts). 2) Excluding very slow phenomena that can have cumulative effects in the long run. 3) Unanticipated changes in the number of thermodynamic phases. 4) Deficiencies in structure of physical correlations and their dependencies on local physical conditions. For example, some species in trace amounts in real physical systems may have been ignored in a model for being small. However, these species may accumulate over time and start impacting key phenomena in an adverse way (e.g., causing deposits on a catalyst surface and inhibiting reactivity). Leaving such species out of a model will obviously cause trouble in applications.

Poorly identified values of model parameters. First-principles models typically require numerous parameters (which may be challenging to measure exactly), such as material properties, initial conditions, or boundary conditions. When lacking sufficient independent experimental data such parameters are hard to estimate, often resulting in unrealistic interdependencies between model parameters.

Computational demands. Detailed FPMs, especially for large or highly detailed systems, demand substantial computational resources. FPM software packages (e.g., Aspen Plus, gPROMS, etc.) are executed in one computing core, and if the model is very large (e.g., model of a large process plant developed for optimization) a dedicated server is used to run the model for efficient and productive use, with extra costs.

Reliability of FPM in transient mode. FPM in transient conditions (start-up, shutdown, transition from one steady state to the other, etc.) are less reliable than conventional steady state mode. In these dynamic models the trends of variations in the variables (e.g., temperature, pressure) are generally correct, but the actual values can be off from reality meaning that such models cannot be used to analyze the dynamics of a plant.

Trust of industrial stakeholders. Although FPM have been used for years and nowadays they are quite advanced, some industrial stakeholders (e.g., operators and maintenance) still do not rely on these models even in steady state mode, which makes it difficult to use them in some areas in the chemical industry.

Fortunately, progress in computational methods and algorithms has helped mitigate some of the limitations by:

Automated model generation, using techniques such as symbolic computation, machine learning, and data-driven modeling that can assist in deriving models from foundational principles efficiently. These methods can help identify relevant mechanisms and simplify complex models without losing essential dynamics.[42]

Parameter estimation using advanced algorithms, including Bayesian inference, optimization techniques, and machine learning approaches, which can improve the accuracy of parameter identification.[20,43]

Data assimilation by integrating real-time data into models (e.g., Kalman filters or variational techniques), thus enhancing their accuracy and predictive capabilities, reducing uncertainties and computational costs.[44]

3. DATA-DRIVEN MODELS

Using only data-driven models in chemical engineering (based on machine learning and neural networks) can offer significant advantages, such as rapid development for interpolation tasks, exceptionally fast prediction times, and the powerful ability to learn complex, nonanalytical patterns and relationships directly from high-dimensional data without prior mechanistic knowledge.[38] This modeling exploits the architecture of artificial neural networks.[45,46] However, these strengths are counterbalanced by several critical limitations associated with relying solely on data-driven approaches that can restrict their applicability and reliability in chemical engineering contexts. Complex data-driven models can capture behaviors beyond current mechanistic understanding, providing performance gains, but at the cost of reduced explainability. The choice to accept this trade-off between performance gain and understanding depends on application context, regulatory requirements, and the risk tolerance. In some domains, this is clearly the right choice, but in others, the loss of understanding creates unacceptable risk.

This section focuses on the limitations of using only data-driven models. We also assume here that the set of inputs to a model always fully span the input space (which is not always the case in reality), and there are no redundant outputs that are combinations of other (independent) outputs.

3.1. Rationale for Data-Driven Models

Before getting into the limitations of DDM, it is important to present clear use cases where DDM works very well:

Cost-driven adoption. For processes with uncertain longevity or proprietary configurations, developing detailed FPM is cost-wise unjustifiable. DDM provides rapid, disposable models.[47]

Computational acceleration via surrogates, as surrogate modeling of expensive FPMs enables rapid optimization, robust design, and extensive sensitivity analysis without repeatedly solving high-fidelity models.[37,48]

Accuracy enhancement via correction, as DDM can correct FPM deficiencies in specific regimes (e.g., neural network correction layers for kinetics correlations under nonideal conditions).

Modeling phenomena beyond FPM, as for example phenomenological DDM can benefit systems where rigorous mechanistic modeling is intractable (e.g., fouling, complex multiphase behavior).

Data-driven hypothesis generation, as DDM can discover unexpected correlations (e.g., multivariate interactions, nonlinear dependencies) suggesting new mechanistic investigations.

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

Rapid model development for transient needs. Short-lifecycle processes, temporary operational changes, or one-time design studies benefit from rapid DDM development without FPM infrastructure investment.

3.2. Data Availability and Quality

Data-driven models are fundamentally dependent on high-quality, extensive data sets to learn accurate representations of the underlying system.[49] Acquiring sufficiently good quality data can be challenging due to several factors:

- *Sparse data*, as many industrial processes operate under conditions where data collection is limited (or the input data are restricted to limited subspace), either because of operational costs, safety concerns, or experimental constraints. For example, data under extreme or novel process conditions might be unavailable, hindering the model's ability to generalize.
- *Noisy and incomplete data*, as sensors and measurement instruments can introduce noise (which misleads the learning process), and data streams may miss entries, resulting in less accurate or overfitted models.
- *Limited data* for rare events, as critical safety-related events or rare process anomalies are infrequently (or not at all) recorded, yet modeling these situations remains crucial for safety. The scarcity of such data impedes the model's capacity to predict and manage these events.

The dependence on data quality means that in many real-world scenarios, the models' accuracy and robustness are compromised, leading to unreliable predictions, especially when the data does not comprehensively cover the full operational space.[12,50] Moreover, training data is often scarce for new processes or extreme conditions, thus limiting the applicability of DDM and reinforcing the need for FPM in innovation and scale-up of chemical processes.

3.3. Extrapolation Challenges

Machine learning models are typically excellent at interpolation, predicting well within the range of their training data, but often fail miserably when extrapolating beyond it. This poses safety risks, especially when models predict unphysical outcomes.[11] In engineering applications, the challenges include:

Limited generalization, as in case of process conditions change (e.g., new feedstock, other temperature ranges), the model may encounter inputs outside its training domain, leading to inaccurate outputs.

Catastrophic failures, as without physical constraints, the models can produce nonsensical results, such as negative concentrations, unrealistic pressure drops, or violations of mass and energy conservation, which can be dangerous if used for real-time control or safety assessments.

Safety risks, as relying on models that cannot reliably predict outside their training domain poses major safety concerns, especially in chemical industry where using such models under unanticipated conditions could lead to accidents.

Recent advances show that overparameterized neural networks can generalize remarkably well, contradicting classical statistical wisdom.[51] However, this success is limited by several aspects:

Domain specificity, as overparameterized success applies primarily to domains where training data well-represents the operational space (e.g., image classification on ImageNet-like distributions). Performance deteriorates sharply in domains with distribution shift, rare events, or extreme conditions.

Opacity remains. Even if overparameterized NNs generalize, their physical inconsistency and opacity persist. In safety-critical chemical engineering (e.g., runaway reaction prediction, heat exchanger fouling), this opacity is unacceptable regardless of statistical performance.

Theoretical understanding lacking. Overparameterized networks work remains poorly understood theoretically, hindering principled application to new domains and our ability to predict when they might fail catastrophically.

3.4. Physical Law Compliance, Interpretability, and Adaptability

Purely data-driven models do not include explicitly any physical laws, thus risking violations of thermodynamic constraints or mass and energy conservation principles. This can lead to physically impossible predictions and undermine model reliability. Often, changes are deliberately made in a better functioning of the object of modeling over its life span of application. Afterwards, its model needs to be adapted as well.

Violation of thermodynamics and conservation laws, as without embedded physical constraints, models might suggest physically impossible states (e.g., perpetual motion or negative energy), undermining trust.

Lack of interpretability, as black-box models offer limited insight into the underlying processes, making it difficult to diagnose errors or understand system behavior, thus hindering validation and troubleshooting.

Difficulties in model validation, as additional steps are required to verify the outputs of these (physical lawless) models against known principles, thus complicating deployment in safety-critical systems.

Difficulties in adapting or extending a model when the model prototype is changed or its functionality extended during its physical lifetime. There is no way of knowing how to adapt the data-driven model to the new situation. The consequence is making a full effort in reidentifying the adapted model, whenever a new extended data set becomes available.

Restricted usability. Neither can a data-driven model advise or guide internal improvements of its prototype, because the model does not represent the inner workings of its prototype in an insightful, mechanistic way. Hence, a data-driven model cannot innovate its existing prototype. It is restricted to supporting operations within the space that was spanned by the data for model identification.

3.5. Safety and Cost Implications

In safety critical applications (e.g., nuclear reactors, chemical reactors, aerospace) model errors can lead to catastrophic consequences. Wrong (nonphysical) predictions may cause safety hazards, equipment damage, or very costly failures. Thus, without any physical constraints, reliance on black-box ML models is risky due to:

Risk of catastrophic failures, as unphysical predictions can lead to unsafe operating conditions, equipment damage, or environmental hazards.

High costs of errors, as incorrect predictions may cause costly process modifications, shutdowns, or product quality issues.

Regulatory and liability concerns, since regulatory agencies scrutinize such models that do not incorporate physical laws, as they might be deemed unreliable or unsafe for decision-making.

Dependence on data quality and quantity, as the need for extensive, high-quality data increases operational costs and may delay deployment, especially in new or proprietary processes where data is rather scarce.

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

Model blindness. In case of a recursive learning practice with a data-driven model an inherent risk may arise by using biased information, i.e., supposedly "new" external data is used for model training while some of that data may have been generated by earlier uses of the model or its earlier versions. This creates a dangerous feedback loop of "model blindness."

3.6. Examples of Data-Driven Models in Chemical Engineering Practice

The success of control-oriented models reflects that control systems operate within bounded regimes where extrapolation is limited and operator intervention provides safety guardrails. In this domain, the ability of complex data-driven models to capture nonlinear dynamics and interactions without explicit mechanistic representation provides significant practical advantage. However, this success should not be generalized to design or process innovation tasks where extrapolation beyond a limited training domain is essential. For example, a novel reactor design exemplifies applications where a predominantly physics-based approach is required due to the need for significant extrapolation and mechanistic understanding. This is in contrast to control applications where the bounded operating regime permits greater reliance on data-driven components.

A couple of examples are given here, illustrating conditions for successes and failures of data-driven modeling.

Example 1: Soft sensors for product quality estimation. A very common and successful application of DDM in chemical engineering is the development of soft sensors. These are models that infer difficult-to-measure product qualities (e.g., polymer melt index, distillation column end points) from easy-to-measure process variables (such as temperatures, pressures, flows). For instance, Fortuna et al.[52] demonstrated the use of neural networks to reliably estimate product composition in a distillation column, providing real-time monitoring without the need for expensive and delayed online analyzers. The success of such models is highly dependent on the training data covering all relevant operating regimes. If the process is modified or new feedstocks are introduced, the soft sensor can become inaccurate, requiring retraining with new data—which is a key limitation where FPM's adaptability would be superior.

Example 2: Performance indicators of multiphase products and materials. Such products have a continuous phase, which hosts bubbles or droplets or islands of the other phases in a dispersed manner. Each phase in a multiphase product has its own internal physical state (p, T, $\underline{x}$) and its local physical properties are known (viscosity, surface tension, conductivity). Yet, it is hard to predict on theoretical grounds only the physical properties of the bulk of the product and its practical performance characteristics (as indicators for practical product quality by product users). Thus, data-driven modeling can be applied.[53]

Example 3: Predictive maintenance for rotating equipment. Purely data-driven models are extensively used for predictive maintenance of critical equipment, such as pumps, compressors, and turbines. ML models can be trained using historical vibration spectra (i.e., frequency domain data), labeled with equipment health states (such as normal operation, imbalance, bearing fault, or cavitation). By applying techniques, such as convolutional neural networks (CNNs), these models can learn to identify subtle patterns that signal early stage failures well before they escalate into downtime or catastrophic damage.[10,54] This enables the implementation of condition-based maintenance, optimizes the maintenance schedules, and prevents unexpected shutdowns. However, the performance of these models relies heavily on the availability and quality of labeled historical fault data, which can be particularly limited, thus posing a significant challenge for model training and reliability.

Example 4: Anomaly detection for pipeline integrity. Data-driven models are powerful tools for pipeline leak detection and anomaly identification. Instead of relying on a single physical model, systems can use a purely data-driven approach by training models such as autoencoders on historical time-series data (of pressure, flow rate, temperature, and acoustic signals) under normal operating conditions. The model learns the normal pattern of operation. During real-time operation, the model reconstructs the current sensor readings. A significant reconstruction error (anomaly score) indicates a potential leak, blockage, or other integrity breach.[12,49] This method can detect smaller leaks than traditional mass-balance methods but requires extensive, high-fidelity sensor data covering various operational regimes to minimize false alarms.

Counter-example: Scale-up of a novel reactor. The design and scale-up of a novel reactive separation process, such as a dividing wall column with reactive zones,[55] is an area where purely data-driven approaches fail. The initial lack of data for a new configuration makes DDM impossible. Moreover, the nonlinear interactions between reaction kinetics, vapor−liquid equilibrium, and hydrodynamics are way too complex to learn from data alone, without any solid physical foundation. First-principles models, based on conservation laws and constitutive relations, are essential to guide the process design, make reliable scale-up predictions, and maintain operational safety. Of course, data-driven techniques may later be used to refine certain submodels (e.g., a kinetic parameter correlation) within the overall FPM framework, but DDM alone rarely suffices for fundamental process innovation (e.g., novel reaction pathways, catalytic breakthroughs, new/intensified unit operations), though it aids incremental improvements (such as anomaly detection, predictive maintenance, process parameter optimization, or finding presence of phenomena not accounted for in a FPM).

4. ANALYSIS OF MECHANISTIC MODEL BUILDING AND USES

4.1. Scope and Purpose of the Analysis

Any computational engineering model is an artifact representing some selected features of an engineering object. Such object either exists in physical reality, or it is of a conceptual nature. Engineering models are created by design concerning practical aspects of the functioning of the object of modeling. This section presents practical attributes of a chemical engineering model that must be considered in its design. The section continues with an analysis the workflow of building of first-principles models. The purpose of this analysis is to identify options for effectively extending first-principles modeling with some data-driven modeling, in case a first-principles model is not fully satisfactory in its applications. The next subsections will deal with the attributes of a model, the context of a model and the workflow in a modeling effort. Understanding model attributes and workflow is instrumental in identifying options for extending a first-principles model with some data-driven black box parts.

4.2. Context of Modeling and Model Usage

A chemical engineering model relates to a specific object in a chemical engineering system having a physical presence and function there. A model is a quantitative image in the

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

mathematical and computational domains of some selected functional features of its object of reference. These features involve aspects of behavior, structure and performance of the object. Such an object is also called the *prototype* of that model.[6] A prototype exists either in a physical space or in a conceptual form. Chemical engineering objects contain chemical and physical resources such as chemical elements and species, mass, electrical charge, energy, momentum. Apart from its internal physical contents, a chemical engineering object has interactions with its environment. Models account for both physical contents and interactions. Often, only a relevant subset of the interactions needs representation in scenarios for model uses; e.g., a model may represent steady state behavior while its object also exhibits dynamic modes. The practical use of a modeling object establishes purposes for model use, as shown in Figure 2.

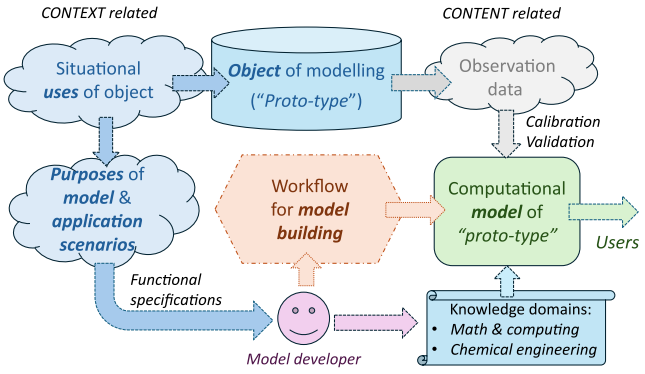


Figure 2. Conceptual environment of a chemical engineering modeling effort.

Chemical engineering objects often serve commercial and/or public functions that impact the economy as well as natural and social environments. Hence, these objects are often rated with respect to relevant performance metrics in the economic, ecological and social domains. In addition to its functional roles, a model is related to its object of modeling by the:

- *Identities of resources* (material, energy, information) *exchanged between object and its environment.*
- *Physical states of these resources* within feasible ranges.
- *Scenarios for changes in the states of external resources and their supply.*
- *Performance metrics and targets* (technological, economic, ecological, ...).
- *Restrictions on the feasible domain of functioning of the object of modeling.*
- *Available data sets* containing observed behavior and performance of the object.

Practical uses of a prototype often deliver sets of measured data. Such data are complementary to existing (a priori) physical knowledge of the prototype and therefore vital to model calibration and validation.

4.3. Practical Attributes of a Chemical Engineering Model

Hereby we suggest a framework of essential attributes to address the fundamental question: *What constitutes a proper chemical engineering model for practical use?* These attributes are not just some academic checkboxes, but key indicators of a model's robustness, fitness for purpose, and the total cost of ownership over its lifecycle. Such attributes offer a systematic foundation for comparing different modeling paradigms, and for developing robust hybrid modeling strategies.

Physical consistency and fidelity: A model must adhere to the fundamental laws of nature. This is the non-negotiable cornerstone of reliable engineering. Key aspects include conservation laws, thermodynamic consistency, causality (correctly represent cause-and-effect relationships inherent in the physical system). FPM inherently excels here, providing a strong foundation. DDM typically lacks this attribute unless explicitly constrained, which is a primary motivation for hybrid modeling.

Interpretability and transparency: A model should provide insight where possible. In this respect, FPM offers mechanistic insight, while DDM reveals hidden data patterns, so both are valuable but still distinct. Notably, insight goes way beyond a specific answer by showing patterns (spatially/time wise) and being able to explain it, even when the insight is restricted to a local domain and not universal. The ability to understand why a model produces a certain output (transparent reasoning) is essential for: debugging and troubleshooting (diagnosing root causes of model errors or plant malfunctions), building trust (acceptance from stakeholders, e.g., operators, managers, and regulators), knowledge discovery (revealing new insights about the system being modeled). FPM excel in this regard as they are inherently transparent, with equations mapping directly to physical mechanisms. However, DDM is often a "black-box," making this a significant challenge. Hybrid models aim to preserve interpretability by keeping the core structure physical.

Validity domain and extrapolation capability: A robust model must clearly define its boundaries and perform reliably within them. This involves explicit demarcation (knowing the ranges of operating conditions, compositions, and scales for which the model is validated), and guarded extrapolation (the ability to make reasonable predictions slightly outside the training/validation domain). This is critical for innovation, scale-up, and handling process upsets. Based on physics, FPM generally extrapolates more reliably. DDM is powerful for interpolation but can fail catastrophically outside its training data. A key benefit of hybrid models is extending the operational range where model is sufficiently accurate.

Operational practicality: A model must be fit for its intended operational role. This includes computational efficiency (solve-time appropriate for its application), numerical robustness (the model should converge reliably to a physically meaningful solution across its validity domain), implementation cost (this includes the total cost of ownership: initial development, validation, maintenance, and required computational resources). FPM can be computationally expensive, while DDM can be very fast once trained but may have high data acquisition and validation costs. Hybrid models often seek an optimal balance, using surrogates to accelerate FPMs.

Adaptability and extensibility: Chemical processes and products evolve, so their models must be able to follow as well. This attribute measures the ease with which a model can be updated to reflect new operating conditions (changes in feedstock or product specifications), new physics (addition of new chemical species, reaction pathway, or unit operation), and new knowledge (improved constitutive relations or property data). Adaptability and extensibility of FPM and DDM differ depending on change scope. For parametric changes (e.g., new feedstock composition within similar range) FPM may require parameter refitting while DDM via transfer learning may adapt faster.[17] In case of structural changes (e.g., new reactive species,

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

Table 1. Key Attributes of First-Principles Models (FPM), Data-Driven Models (DDM), and Hybrid Models

Attribute	Description	FPM	DDM	Hybrid models
Physical consistency	Adherence to conservation laws, thermodynamics, and causality	Inherently high	Typically low (unless constrained)	High (physics-based foundation)
Interpretability	Ability to understand and trace model outputs to underlying mechanisms	Inherently high (white-box)	Typically Low (black-box)	High (transparent core)
Extrapolation capability	Reliability of predictions outside the immediate training/validation domain	Strong	Very weak (high risk)	Enhanced (beyond pure DDM)
Computational efficiency	Speed of model evaluation for its intended use case (e.g., real-time)	Variable (often slow)	High (after training)	Optimized (surrogates for speed)
Development and adaptation cost	Effort required for initial build, validation, and future modification	High initial, low cost adaptation	Low initial, high cost adaptation	Balanced (leverage strengths of both)
Data requirements	Quantity and quality of data needed for model construction/calibration	Moderate (for parameters)	Very high (for training)	Reduced (physics reduces data need)
Uncertainty quantification	Ability to quantify confidence in predictions and identify error sources	Strong framework	Limited (often statistical only)	Robust (combined methods)

additional phenomena) FPM readily accommodates via new equations, but DDM typically requires full retraining. As for extrapolation to new regimes (e.g., higher temperatures, new catalysts) FPM physics provides guidance, but DDM requires extensive new data or struggles with extrapolation. This comparison is context-dependent, not absolute. However, in safety-critical and innovation-intensive applications where structural flexibility and extrapolation capability are paramount, FPM's adaptability remains superior.

Uncertainty quantification (UQ): A reliable model must quantify its own limitations. It should provide not just a prediction, but also a measure of confidence, stemming from parameter uncertainty (propagation of errors from fitted parameters), structural uncertainty (errors arising from missing physics or incorrect model assumptions), and input uncertainty (errors from noisy or biased sensor data). FPM provides a natural framework for uncertainty quantification (via sensitivity analysis), but it faces some challenges. Structural uncertainty (unknown phenomena, missing species, and incorrect assumptions) is difficult to quantify. Parameter uncertainty propagation assumes the model structure is correct, which may not hold. DDM can estimate prediction variance but often struggles to separate epistemic (model structure) from aleatoric (data noise) uncertainty. Among challenges, the epistemic uncertainty (model inadequacy for out-of-distribution inputs) versus aleatoric uncertainty (data noise) is quite difficult to disentangle. Ensemble methods and Bayesian approaches (e.g., Bayesian NNs) improve UQ but remain computationally expensive. As neither FPM nor DDM has a clear UQ advantage, hybrid approaches would be recommended. Physics-informed Bayesian neural networks (PBNNs), ensemble hybrid models, and constrained uncertainty propagation show promise in addressing both FPM and DDM limitations.[56]

This framework of attributes provides a valuable scorecard against which any model can be evaluated for a specific context of use. The ideal model maximizes the desired attributes for its application. In practice, this often leads to a hybrid strategy: using FPM to ensure physical consistency, interpretability, and adaptability, and judiciously integrating DDM to enhance computational efficiency, refine predictions within the validity domain, or model phenomena where first-principles are incomplete. A comparison of FPM, DDM and hybrid models with respect to these attributes is conveniently summarized in Table 1.

4.4. Building the Contents of a Model

The life span of a model comprises three distinctive phases: 1. Initial model development and validation; 2. Use of a model when making applications; and 3. Adaptations/extensions of a model for wider applications. The initial development of a model requires (like any other engineering effort) a lot of rational decision making, concerning the function, structure and behavior of the model, while matching the goals with the available means. The initial model building is often done by means of a multistep structured workflow (Figure 3).

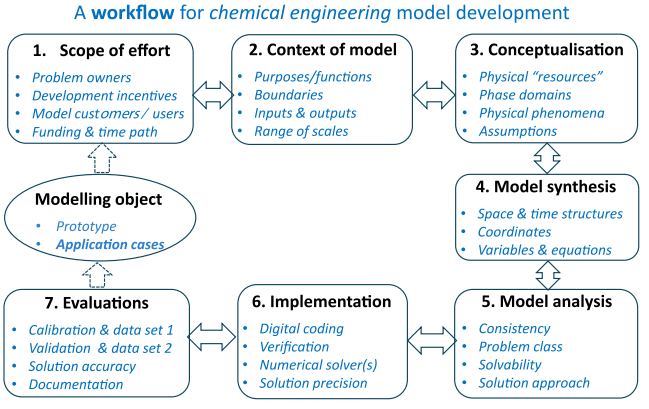


Figure 3. Workflow for the development of a mechanistic chemical engineering model.

Remarkably, only a fraction of academic textbooks on chemical engineering modeling offers explicitly a workflow of modeling, before addressing the necessary mathematical and computational technicalities of modeling one by one.[8,22,28]

Structural contents: Having established the context of model uses, the next phases of model building involve the creation of contents of the model in its mathematical and digital forms. This takes place in four steps: conceptualization, synthesis and analysis and implementation (steps 3−6 in Figure 3). The conceptualization of a model is the more crucial one. This step defines both the structure and size of the mathematical model as well as the features of its mathematical behavior. The structure and size of a model are the result of many decisions of a discrete nature (yes or no) with respect to the presence of physical items in the object of modeling, and hence in the model. A list of assumptions is made about the presence of spatial domains and the physical resources and physical phase (s) per domain. Each entity per domain will induce either an additional equation, or an additional term in a resource balance equation. These structural decision items apply to each processing node in a processing network. There is a distinct multiplication effect with respect to the number of entities being considered in a model. Table 2

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

Table 2. Discrete Structural Decisions in First-Principles Model Building and Opportunities for Hybrid AI/Data-Driven Integration

Category	Decision item	Description/implication	Opportunity for hybrid/DDM approach
Spatiotemporal resolution	Spatial dimensions and discretization	Choice of 1D, 2D, or 3D model; fineness of mesh. Impacts computational cost vs accuracy.	AI-generated adaptive meshing. ML-surrogates for high-fidelity subdomains.
Spatiotemporal resolution	Dynamic vs steady-state	Whether to model transient behavior. Essential for control and safety, but more complex.	AI for model order reduction of dynamic models; real-time data assimilation (e.g., Kalman filters).
Governing principles	Conservation laws	Selection of which laws to enforce (e.g., mass, energy, momentum, charge). Non-negotiable core.	PINNs to embed these laws as hard constraints in otherwise data-driven components.
Governing principles	Thermodynamic framework	Choice of equation of state (EOS) or activity model (e.g., ideal, NRTL, SAFT). Major source of uncertainty.	ML-corrections to simple EOS; NN-based property predictions replacing complex correlations.
System constituents	Chemical species inclusion	Decision to include/ignore trace components or side products. Can be critical for safety and fouling.	Data-driven anomaly detection to identify unmodelled species affecting performance.
System constituents	Phases present	Assumption of number and type of phases (vapor, liquid, solid). Phase changes are complex.	ML classifiers to predict phase boundaries or transitions from operational data.
Phenomena and kinetics	Reaction mechanism and kinetics	Postulation of reaction pathways and rate expressions. Often the largest knowledge gap.	Prime candidate: Replace empirical rate expressions with NN-learned functions from data.
Phenomena and kinetics	Transport phenomena	Modeling of mass/heat transfer (e.g., diffusion coefficients, transfer rates). Often uses semiempirical correlations.	Prime candidate: Data-driven refinement of mass/heat transfer coefficients within the FPM framework.
Model boundaries and purpose	System boundary and lumping	Defining the unit operation or process scope. Lumping components or units reduces complexity.	Data-driven models to represent lumped subsystems or entire supply chain networks.
Model boundaries and purpose	Key Performance Indicators (KPIs)	Selection of model outputs relevant to the purpose (e.g., purity, yield, energy use).	AI/ML for rapid optimization and sensitivity analysis around these KPIs using the FPM.

presents an overview of discrete structural decisions in first-principles model building and opportunities for hybrid models integration.

Synthesis of model equation sets and variables: Having the structure per region in place one can introduce the associated mathematical variables and equations per spatial region as well as for the transfers of resources between regions. While the structure of the resource balance equations is predetermined, there will often be choices with respect to the nature of empirical correlations for physical properties, transfer and rate phenomena. Each region will have its own physical state (vector $\underline{x}$). Causality of events is a key feature of first-principles modeling, implying that the physical properties depend on the local state while the rates of local phenomena depend on gradients in the local physical state.

4.5. Model Implementation and Numerical Solution

Implementation: The transition from a conceptual mathematical model to a robust, computable digital implementation is a critical and often underappreciated phase in the modeling workflow. This step involves digitally coding the model, verifying the model implementation for correctness, and linking it to suitable numerical solution techniques.[29] The choices made here directly impact the model's computational efficiency, numerical robustness, and ultimately, its practical utility. This section details the key technical considerations and best practices for successful model implementation. The first step for a proper implementation of a model is equation formulation and scaling. The equations are put in a form that complies with the required format of the chosen numerical solver. To enhance numerical conditioning of the model often scaling is applied to variables and equations.

Solver Selection and Configuration: Selecting the appropriate numerical solver is paramount and depends on the mathematical problem class (algebraic, differential, partial differential and/or integral equations). Within a problem class and depending on model's characteristics (e.g., stiffness of model) one can opt for the most suitable solver within that class.

Verification via Trivial and Analytical Cases: Verification ensures the computational model accurately represents the intended mathematical model. This involves testing with trivial cases (e.g., checking conservation laws sum to zero) and against any available analytical solutions for simplified scenarios.[33,34] Before attempting to solve complex scenarios, the implementation must be rigorously tested by 1) *Conservation checks*: run simulations to test the conservation of elements, mass, energy and electrical charge. 2) *Simple case testing*: test the model under simplified conditions where behavior is known. 3) *Comparison with cases having analytical solutions.*

Solving a Model and Addressing Numerical Challenges: These may involve the initialization of an iterative model computation; handling of discontinuities caused by external or internal model events; high sensitivities of model outcomes to parametric changes.

By properly addressing these aspects, a modeler can transform a theoretical construct into a reliable digital asset. This robust foundation is a prerequisite for the subsequent steps of parameter estimation and validation, ensuring that any discrepancies discovered are due to model structure or parameters, and not numerical artifacts.

4.6. Model Evaluations

Model evaluations involve parameter estimation and model calibration. The latter are intertwined processes of refining a mathematical model by adjusting its uncertain parameters to minimize the discrepancies between model outputs and real-world observations. Model parameters in constitutive equations are estimated from experimental or industrial data. Sufficient spread in data is needed for identifying parameters in mechanistic models to cover the full output space evenly and to avoid empty subspaces, ensuring good identifiability. Parameter estimation is typically formulated as an optimization problem minimizing the error between model predictions and data, while accounting for the probabilistic nature of measurement errors.[20]

Model Calibration. Requires that the computed solutions by the model for a specified range of scenarios must agree with independent external data for these scenarios. It may require tuning some model parameters, or making changes to model contents. Upon agreement between computed model data and

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

the external reference data, a model is validated and ready for the next phase of making routine applications.

Life Span Considerations. Often a model will be adapted over its lifetime, depending on changes in functionality of the external object it represents. A critical reflection involves how well available calibration data spans the input and output spaces. If the data occupies only a subspace, the model's predictive reliability is compromised outside that subspace. This is fundamental limitation to both FPM calibration and DDM training. It underlines the risk of extrapolation and highlights a need for physical constraints to prevent unphysical predictions.

4.7. Extending a First-Principles Model to a Hybrid Model

Building a first-principles model is based on a body of prior mechanistic knowledge. However, uncertainties may arise in models due to lack of knowledge of the nature and behavior of chemical and physical phenomena. For instance, the active presence of a physical resource in a tiny amount may have been overlooked. Hence, there is no upfront guarantee that modeling decisions will result in a model that properly captures all relevant behavioral features of its object. Testing of models with empirical data is needed before these can be put into practical use, for which two complementary options exist:[8,35]

1 *Estimation of the model parameters* by means of available empirical data on the behavior and performance of the object. Such estimation is part of a wider model calibration effort. In case the estimation results are not satisfactory by leaving a significant gap between model outputs and data, one could consider making changes in the structure of the model as part of a wider model calibration effort.
2 *Validation of a model* by means of independent sets of data; that is, independent of the data sets used for parameter estimation. If its validation outcome is satisfactory, a model can be accepted for routine use in its domain of application. Otherwise, model adaptations must be considered beyond the mechanistic domain.

Persistent mismatches remain between model and data can be due to unknown structural physical deficiencies in the model. For instance, by ignoring the presence of an active physical resource, such as a reactive chemical trace component. Or some unknown physical phenomena may play out, such as the unanticipated splitting of a thermodynamic phase in two phases. In those situations, a pragmatic link with data-driven modeling comes into view, combining a first-principles model selectively with some black-box components. For example, Bradley et al.[11] present six structural options for combining the data-driven and first-principles parts into a hybrid model; Zhou et al.[53] apply hybrid modeling for multiscale integrated material and process design; Zuo et al.[57] create a deep learning hybrid framework for modeling actual distillation processes; while Gargalo et al.[58] reviewed the use of hybrid modeling in the (bio)chemical industries in the context of Industry 4.0.

5. INTEGRATED FIRST-PRINCIPLES AND DATA-DRIVEN APPROACHES

Integrating FPM and DDM approaches is essential to take synergistic advantage of both engineering principles and data-science at the same time.[59] This is the case in various situations, such as in the correction of a simple mechanistic model that does not fully capture the physics of the system, the integration of a data-driven component that approximates an unknown correlation within a mechanistic model, or just in the development of a faster model of an accurate high-fidelity model (also known as surrogate models).

Different techniques have been proposed and applied to achieve this integration, such as hybrid modeling, physics-informed machine-learning and model calibration, as reviewed by Bradley et al.[11] and Chakraborty et al.,[60] who provided a comprehensive comparison of the hybridization techniques with respect to their differences and similarities, as well as advantages and limitations. Rajulapati et al.[61] identified and categorized various hybrid models that integrate ML models with various forms of domain knowledge, leading to benefits such as enhanced predictive power and extrapolation capabilities. A recent review paper provides a step-by-step tutorial review for practical implementation of hybrid modeling, combining FPM with ML.[62]

Nowadays, AI and industrial data can be embedded into process simulators to enhance the model predictions and develop accurate hybrid models. One can benefit from integration with existing process simulators such as Aspen HYSYS and Aspen Plus. For example, taking an existing simulation built in Aspen Plus and transforming it into an Aspen Hybrid Model can leverage the power of AI without engineers requiring data science or machine learning expertise.[63]

While tools such as Aspen Hybrid Models offer API access and DDM-FPM coupling capabilities, real-world integration faces several well-documented challenges. Not surprisingly, some industrial cases[64] reported 30−50% longer deployment times than vendor claims.

- *Data pipeline mismatches*. Real-time data streams from plant sensors often operate at different time scales and formats than batch simulation environments used for FPM solving. Reconciliation requires complex data curation and transformation layers.
- *Loss of interpretability*. When embedding black-box DDM components within transparent FPM structures, the resulting hybrid loses some interpretability benefits. Tracing errors or understanding predictions becomes more difficult.
- *Version control and maintenance complexity*. Hybrid models require version management of both FPM code and DDM model parameters/weights simultaneously, often across different software ecosystems and organizations, creating deployment and troubleshooting bottlenecks.

5.1. The Rationale for Hybrid Models

The rationale for hybrid models stems from the recognition that neither purely physics-based models, nor purely data-driven models alone, can fully address the complexities of many real-world systems.[58] For instance, in almost any mechanistic model there will be a cutoff scale in modeling detail below which one needs a different approach. Then one can use a more detailed model for the tinier scales, such as a statistical particle population model instead of a continuum description. Or one can resort to a data-driven description, such as empirical correlations that link thermodynamic forces and fluxes by means of empirical transfer coefficients (for mass, energy, momentum, electrical charge, etc.). The key fundamental requirements for hybrid approaches are:

1 Maintaining the consistency of metrics in time and space.
2 Conservation of physical entities being conserved in nature. This involves number of atoms of a chemical

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

element, mass, energy, momentum, electrical charge, number of electrons, ...

3 Non-negative entropy production rates (second law of thermodynamics), and

4 Respecting causality, namely cause-effect sequences.

Combining FPM with ML in an effective manner enables hybrid models that can leverage the physics-based understanding with data-driven flexibility. Such an integrated approach results in models that are more robust, accurate, and transparent,[11,65] thus enhancing their applicability across real-world problems in many fields, such as engineering, environmental science, and biomedical research.

5.2. Hybrid Modeling Techniques

The integration of FPM and DDM is not a uniform process but rather encompasses a range of methodologies, each offering distinct benefits and suitable applications. Looking beyond the example of Physics-Informed Neural Networks (PINNs), these approaches can be classified according to the function of the data-driven element within the physics-based structure. This taxonomy helps identifying the most appropriate integration strategy for a given modeling challenge. Figure 4 illustrates some of the hybrid modeling techniques.[66] The choice of technique depends on the specific modeling context, as follows (key examples only):

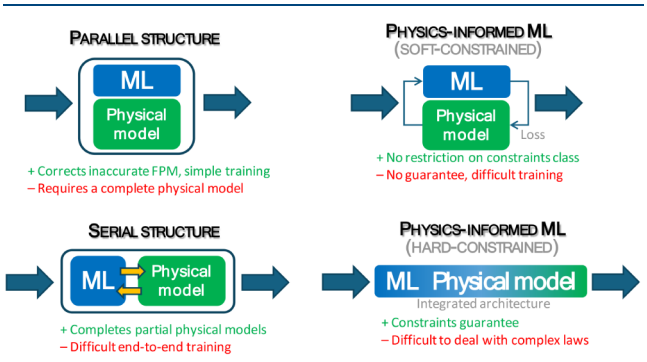


Figure 4. Hybrid modeling techniques: parallel structure, serial structure, and physics informed ML (soft/hard-constrained).

***Physics-informed learning*(*Embedded constraint*).** This approach, exemplified by PINNs, takes a rather direct approach to integrating physics into machine learning, as they actually embed the governing equations right into the loss function of the learning algorithm.[65,67] In practice, this means the neural network must not only fit the available data, but also minimize the residuals of physical equations (e.g., conservation laws or reaction kinetics) across the whole domain. This method is particularly useful for tackling forward and inverse problems involving partial differential equations, especially in situations where data is limited but the underlying physical principles are well understood, as for example in computational fluid dynamics or reaction-diffusion systems. By structuring the learning process this way, the neural network can produce solutions that are consistent with the underlying physics, even when data is scarce. Yet, the main challenge is the computational cost of training and ensuring the network converges to a solution that complies with the physics and fits the data. PINNs have been successfully used to model flow dynamics, heat transfer, and reactive systems with limited data. One should use physics-informed learning when governing equations are known but analytical/numerical solutions are challenging, and data is sparse.

***Serial hybrid modeling*(*Corrector or preprocessor*).** Gray-box hybrid models embed DDM components within FPM structure, where learned functions (neural networks, Gaussian processes) replace unknown constitutive relations or closure models.[68] In such a serial configuration, the data-driven component acts as a distinct module either before or after the FPM. A common application is using a machine learning model (e.g., Gaussian Process or ANN) to predict difficult-to-measure but critical input parameters for the FPM, such as thermodynamic properties or reaction rates (kinetic constants) expressions that directly parametrize FPM mass balance equations. Conversely, a DDM can act as a "corrector" placed after the FPM, postprocessing its outputs to enhance accuracy based on historical errors. This structure maintains a clear separation between physics and data-driven correction, aiding interpretability. One should use serial hybrid modeling when the primary uncertainty is in specific input parameters or a final calibration step.

***Parallel hybrid modeling*(*Residual learner*).** This is one of the most industrially prevalent hybrid forms.[69] Here, a relatively simple or approximate FPM provides a baseline prediction. A DDM (often a NN) is then deployed in parallel to learn the residual (the systematic error or discrepancy between the FPM output and the actual plant data). The final model output is the sum of the FPM prediction and the DDM-calculated residual. This approach powerfully leverages the FPM's extrapolation capabilities while using data to capture complex and/or poorly understood phenomena that are difficult to model mechanistically. It effectively "*learns what we don't know*" while being based on "*what we do know*". One should use parallel hybrid modeling when a mechanistically sound but incomplete FPM exists, and abundant operational data is available to capture the gap. Note that a critical concern in parallel hybrid structures is that when a DDM component runs alongside an FPM component, the DDM can introduce physically inconsistent corrections outside its validity domain. For example, if an NN was trained on data at 50−150 °C and suddenly operates at 200 °C, it may still produce a numerical prediction (extrapolation), but this prediction is unreliable and can physically contradict the FPM's physically grounded prediction. Or if the combined model's output (typically a weighted average or gating mechanism) may mask this inconsistency, leading to silently incorrect predictions. As mitigation strategy one could use domain-aware hybrid control involving:[56] explicit domain tracking (continuously monitor input variables against DDM training bounds); validity warning system (flag when DDM operates outside its domain; revert to FPM-only predictions or reduce DDM weighting); adaptive gating (use a learned gating function that dynamically down-weights DDM contributions outside its domain); and uncertainty inflation (increase reported prediction uncertainty when DDM domain is violated). This risk-aware approach is essential for safety-critical applications.

***FPM-constrained surrogates*(*Accelerator*).** For complex, high-fidelity FPMs that are computationally too expensive for real-time optimization or uncertainty quantification, a common strategy is to build a fast data-driven surrogate (or reduced-order model). The key differentiator of a hybrid surrogate is that it is not a pure black-box, being trained on a vast data set generated by the high-fidelity FPM, but its architecture or training process is constrained to respect fundamental physics.[11,60] For instance, the surrogate's outputs can be constrained to obey mass

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

balances, or its inputs and outputs can be designed to represent physically meaningful state variables. This results in a model that is both fast and reliable, forming the core of many modern "digital twins".[63] One should use FPM-constrained surrogates when a high-fidelity FPM is available, but it is too slow for its intended application (e.g., real-time control, multiscenario optimization).

Optimization-based approaches. In mixed-integer optimization formulations,[70] DDM can predict constraints or uncertainty bounds that are fed into FPM-based optimization problems. This hybrid structure is particularly powerful for large-scale systems where full FPM simulation is computationally prohibitive, but FPM-based optimization is feasible if constraints/objectives are approximated via fast DDM surrogates. Such approaches combine the physical rigor of FPM optimization with the computational efficiency of DDM-based constraint approximation.

Training difficulties in hybrid models (mentioned also in Figure 4) include:

- *Competing loss terms.* FPM demands conservation law satisfaction (hard constraints via physics loss) while DDM maximizes prediction accuracy. These objectives can conflict, creating Pareto frontiers where no single solution satisfies both perfectly.
- *Gradient conflicts.* Back-propagation through coupled FPM-DDM chains can lead to contradictory gradient signals when physics and data-driven losses pull in opposite directions.
- *Sampling bias.* If training data is biased toward certain operational regimes, the DDM component overfits to these regions, degrading hybrid performance in underrepresented scenarios.

However, there are some solutions to tackle these difficulties, as proposed in literature:[11,71]

- *Adversarial cotraining*: FPM and DDM components compete, with the FPM component detecting where DDM violates physics and vice versa, iteratively improving both.
- *Decoupled training*: Train FPM and DDM on sequential phases with adaptive weighting schedules rather than joint optimization.
- *Physics-guided regularization*: Add soft constraints (penalty terms) rather than hard constraints, allowing some flexibility while maintaining physical grounding.
- *Importance sampling*: Use guided data selection techniques to ensure training samples are representative across the full operational space, not just historical regions.

5.3. Residual Learning and Model Refinement

Residual learning and model refinement are powerful strategies in modern modeling approaches, particularly when combining data-driven techniques with traditional physics-based models. Residual learning involves training a model to predict the difference (or residual) between an initial estimate (often derived from a physics-based or FPM) and the actual observed data. This approach allows the residual model to specifically address the errors or discrepancies that the initial model does not capture. By leveraging the strengths of both physics-based modeling and data-driven techniques, residual learning supports efficient and physically consistent model refinement, leading to improved accuracy. Furthermore, this method often reduces data requirements while enhancing overall model performance.[71]

5.4. Benefits and Challenges

Integrating FPM with DDM enhances predictive accuracy, especially in extrapolation regimes, and improves transparency and interpretability. The main benefits of this approach include:

Enhanced predictive accuracy, as hybrid models leverage the strengths of physics-based models to enforce known physical laws, while data-driven components capture the complex, nonlinear phenomena that are difficult to model explicitly. This synergy improves the model's ability to make accurate predictions outside the range of the training data, particularly where pure data-driven models may struggle or fail.

Improved interpretability, as including first-principles ensures the model adheres to established physical laws. This constrains the model's structure, making its behavior more transparent and interpretable vs ML.

Better generalization and extrapolation capabilities, as by embedding physical constraints, hybrid models are less prone to overfitting and can generalize better to unseen conditions, especially in regimes where data are sparse, but physics is well-understood.

Reduction of data requirements, as first-principles provide a solid foundation, reducing the dependence on large data sets and enabling effective modeling even with limited data.

However, there are also some significant challenges such as:

Increased computational complexity, as hybrid models often combine complex physics simulations with data-driven algorithms such as neural networks. This integration can lead to much larger computational costs, thus requiring efficient algorithms, and high-performance computing resources.

Ensuring stability and robustness, as including data-driven components into physics-based frameworks can lead to numerical instabilities, especially if the learned components violate physical principles or introduce discontinuities. Developing stable algorithms that maintain the systems' fidelity is an ongoing challenge. Awareness is required about possible occurrences of solutions based on spurious roots in a hybrid model.

Maintaining conservation laws and physical consistency, as ensuring that the data-driven parts do not violate fundamental laws requires careful design, such as embedding constraints directly into the training process or model architecture.

Algorithm development and training difficulties, as it is nontrivial to develop algorithms that integrate physics-based constraints into training procedures. It often involves complex loss functions, regularization techniques, or specialized architectures, which can complicate the model development and tuning.

Balancing model complexity and interpretability, as increasing complexity may inadvertently reduce its transparency. Striking the right balance between complexity and transparency is an active area of research.

Safeguarding the proper use of a hybrid model, by training its users as well, other than the original model developers.

Complexity of the model for the industrial users, who do not have the required technical background knowledge, will not use these models actively and efficiently so these models will lose their added value in industrial applications.

5.5. Example Applications of Hybrid Modeling

These concise examples demonstrate the use of hybrid principles, showcasing the versatility of hybrid paradigm based

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

on the synergy that yields superior models (as compared to what either approach could achieve alone). The integration strategy is tailored to the specific challenge (e.g., data scarcity, computational cost, or knowledge gaps), and is implemented as a constraint (Case 1), a parallel residual learner (Case 2), a fast surrogate (Case 3), a deeply embedded physics-informed solver (Case 4), or as a FPM-constrained surrogate for system optimization (Case 5). Clearly, extended case studies with detailed model development timelines, computational comparisons, and industrial implementation metrics would exceed the scope of this perspective paper. For readers seeking deeper technical treatments, we direct them to the cited primary sources.

Although most of the DDM examples are related to ANNs, it should be noted that AI encompasses a much broader ecosystem beyond ANNs, including: reinforcement learning (RL), Bayesian probabilistic models, support vector machines (SVM), Gaussian processes, Bayesian optimization, genetic algorithms, and ensemble methods. However, the primary focus on ANN-based examples is justified by their current dominance in industrial AI/ML applications (as most commercially relevant methods), and their prominence in the Generative AI revolution. Nonetheless, non-ANN DDM techniques have their own specific strengths, e.g., Gaussian processes for uncertainty quantification, SVM for high-dimensional classification, Bayesian optimization for expensive-to-evaluate objectives.[42]

Example application 1: Soft sensor with physical constraints. A common application is enhancing soft sensors for product quality estimation. A purely data-driven neural network can predict a product concentration from temperatures and flows. A hybrid approach would ensure the predictions adhere to mass balance constraints, drastically improving reliability during process upsets or when extrapolating slightly beyond the training data. In a polyethylene reactor, a purely data-driven neural network was developed to predict the melt index (a critical quality parameter) from easily measured temperatures and pressures.[52] While accurate under normal operations, this model could produce unphysical predictions (e.g., violating overall mass balance) during process upsets or grade transitions. A hybrid approach was implemented by embedding the neural network within a simplified first-principles model that enforced mass and energy conservation laws as hard constraints during the training process. This constrained learning ensured all predictions were physically plausible. The result was a 40% improvement in prediction reliability during transient operations, along with a major reduction in off-spec product. This demonstrates that physical constraints act as a regularizer, drastically improving the robustness and reliability of data-driven models in safety-critical monitoring applications. To summarize, the DDM limitation is that purely data-driven soft sensors can produce unphysical predictions during process upsets. The FPM limitation is that creating a full first-principles soft sensor can be unnecessarily complex for this application. To address these limitations, a hybrid approach embeds the neural network within a simplified FPM that enforces mass and energy conservation as hard constraints.

Example application 2: Reactor modeling with learned kinetics. The kinetics of a complex catalytic reaction may be poorly known. A hybrid model can be constructed with first-principles mass and energy balances for the reactor, but with the reaction rate term represented by a neural network trained on experimental data. This provides the extrapolation power and physical consistency of the FPM with the accuracy of a DDM for the most uncertain part of the model. Biochemical processes are notorious for complex, time-varying kinetics that are difficult to model mechanistically. von Stosch et al.[69] detailed a seminal hybrid approach for a fed-batch bioreactor. The core model consisted of first-principles dynamic mass balances for biomass, substrate, and product. However, the complex specific growth rate term, $\mu(X,S,t)$, which is a function of biomass (X), substrate (S), and time (t), was replaced with an artificial neural network (ANN). The ANN was trained on historical operational data to learn the relationship between the state variables and the growth rate. This architecture provided the extrapolation power and structural integrity of the differential balance equations while leveraging the data-driven component's ability to capture the intricate, nonmechanistic kinetics. This approach led to a more accurate prediction of the optimal feeding strategy, increasing final product concentration by over 15% compared to strategies derived from traditional, imperfect kinetic models. To summarize, the FPM limitation is that the kinetics of a complex catalytic reaction may be poorly known and difficult to model. The DDM limitation is that a black-box model of an entire reactor lacks physical consistency and extrapolation ability. To solve these issues, a hybrid model uses first-principles mass and energy balances but represents the uncertain kinetic term with a neural network trained on operational data.

Example application 3: Digital twin for real-time optimization. A high-fidelity FPM of a distillation column can be used to generate a vast amount of simulation data across various operating conditions. A physics-informed surrogate model is then trained on this data to create a digital twin that executes in milliseconds. This twin can be used for real-time optimization, providing optimal set points that respect all physical constraints of the system, something a pure DDM could not guarantee. High-fidelity first-principles models of complex units, such as crude distillation columns (CDUs) are computationally expensive, making them unsuitable for real-time optimization (RTO). A solution involved creating a physics-constrained surrogate model.[11,63] A rigorous FPM of the CDU was used to generate a vast data set of input-output relationships across a wide range of operating conditions (feed composition, flow rates, column pressures). A machine learning model (e.g., a gradient-boosted tree or a neural network) was then trained on this data to create a fast-executing digital twin. Crucially, the surrogate's architecture was designed to ensure predictions respected key physical constraints, such as component mass balances and thermodynamic consistency. This hybrid digital twin could execute in milliseconds (instead of minutes), enabling closed-loop RTO. Industrial reports indicate that such implementations can yield a 2−5% increase in margin for a typical CDU by providing optimal set points that respect all physical constraints of the system—a guarantee a pure black-box DDM could not provide. In this example, the FPM limitation is that a high-fidelity first-principles model is often too computationally expensive for RTO. As for the DDM limitation, a pure data-driven model trained on plant data cannot guarantee physical consistency and is unsafe for optimization. The elegant hybrid solution is that a physics-constrained surrogate model is trained on data generated by the high-fidelity FPM, resulting in a millisecond-fast digital twin that respects key physical constraints.

Example application 4: Discovering flow dynamics in chemical vapor deposition (CVD). This example demonstrates a more advanced PINN application. Modeling CVD reactors

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

involves solving complex CFD equations coupled with reaction kinetics—a computationally prohibitive task for many design scenarios. Raissi et al.[65] applied a PINN to learn the flow and concentration fields inside a CVD reactor. The neural network was trained not only on sparse experimental sensor data (for velocity and concentration) but also to minimize the residual of the underlying Navier−Stokes and species conservation equations at a large number of randomly chosen points in the domain. This approach successfully reconstructed the full flow field with high accuracy using 100× less data than a comparable pure CFD simulation would require for grid generation. It effectively "learned" the physics of the system, providing a powerful tool for rapid design exploration and inverse problems where the goal is to deduce operating conditions that yield a desired film deposition profile. To summarize, the FPM limitation is that solving complex CFD equations coupled with reaction kinetics is computationally prohibitive for rapid design exploration. Conversely, the DDM limitation is that a pure black-box approach would require an infeasible amount of experimental data. As a hybrid solution, a PINN is trained on sparse sensor data while also minimizing the residual of the governing Navier-Stokes equations, effectively learning the physics of the system with minimal data.

Example application 5: Dynamic market-driven production optimization in refineries. This example showcases how hybrid modeling enables refineries to adapt plant operation to market demand. The aim is to change the plant production in a reasonably short time (e.g., less than a day) to produce more profitable market-demanded products (e.g., for periods of days or weeks). This practice (operational planning or real-time optimization), can integrate FPM with plant data to rapidly validate and execute optimal operational changes. Refineries are complex, interconnected networks where changing the operating conditions of one unit affects downstream units and the final product yields and qualities. A purely FPM of an entire refinery is computationally prohibitive for frequent reoptimization required by volatile markets. Conversely, purely DDM lacks the physical consistency to safely optimize across wide operating ranges and cannot reliably predict the effects of large changes never before seen in historical data. A hybrid solution involves creating a hierarchy of models. At the core, rigorous high-fidelity FPMs of individual process units (e.g., developed in Aspen Plus or PetroSIM) are used to generate accurate data across a wide range of operating conditions.[64,72] This data is then used to train physics-constrained surrogate models (e.g., neural networks or gradient-boosted trees) for each unit, which are optimized to execute in seconds instead of hours.[48] These fast surrogates are integrated into a refinery-wide multiperiod optimization framework that respects physical constraints (mass balances, product quality specifications and equipment limitations). The model can rapidly calculate the new optimal set points for units across the refinery, in order to maximize margin while ensuring that all products remain on-spec. The model's physical foundation provides the confidence to implement these changes, knowing that predictions of product qualities and yields are reliable. This approach turns optimization from a monthly (or weekly) exercise to a daily (or even shift-wise) activity. Industrial reports indicate that such hybrid RTO systems can yield a large increase in operating margin. For example, such an implementation can lead to 1−3% increase in margin for a typical refinery (equivalent to tens of millions of dollars annually for a large refinery), by dynamically aligning production with market prices, reducing off-spec product, and minimizing energy use during transitions.[64] To summarize, the FPM limitation is that rigorous refinery-wide models are too slow for the frequent reoptimization required by dynamic markets. As for the DDM limitation, a pure black-box model trained on historical data cannot be trusted for large, unseen setpoint changes and lacks built-in physical safeguards. The hybrid solution uses physics-constrained surrogate models of individual units (trained on FPM-generated data) that are integrated into a fast, plant-wide optimization framework. This construction enables safe, rapid, and reliable optimization in response to market fluctuations.

5.6. AI-Assisted FPM Development

Currently, no digital representation alternative to mathematical equations and variables has proven equally effective for capturing quantitative physical behavior with the same extrapolation reliability. Cellular automata, neural networks with symbolic regression, and graph neural networks represent interesting research directions, but none have matched mathematical FPMs. However, we want to introduce here an intriguing possibility: *could Generative AI tools be trained to automatically generate FPMs?* Such a tool could be trained on extensive documentation of validated FPMs, where training data includes: structural representations of systems; model purposes and application scenarios; textually stated modeling assumptions concerning the presence of physical resources, phases and the occurrence of physical phenomena and constitutive relations; sets of equations and parameters; domains of validity. Given this rich, structured training data on mathematical FPMs (which is much smaller and more precisely structured than natural language) it is conceivable that Generative AI could learn to generate new FPM formulations for novel situations. This would represent a paradigm shift: not replacing FPM with AI, but using AI to accelerate FPM creation, thus dramatically reducing development timelines and costs.

Consider for example LLM assistance in generating kinetic rate laws. An LLM could be prompted with mechanism descriptions, literature examples, and chemical principles to generate candidate rate expressions. However, these candidates must be evaluated (using fundamental kinetic and thermodynamic principles and experimental data) to verify thermodynamic consistency, temperature dependence, composition dependencies, and adherence to elementary steps. The LLM provides a starting hypothesis, while the expert human chemists and engineers provide validation. Notably, the LLM-assisted model generation could dramatically accelerate the initial FPM development phase, but the validation phase becomes more critical and requires deeper expertise.

To offer a more concrete vision (or research direction) about an AI-Assisted FPM modeling tool, we propose developing a Generative AI tool specifically trained for chemical engineering FPM development. Such a tool would accept contextual specifications (system structure, phenomena to model, desired outputs, validation constraints) and generate validated FPM formulations. The training infrastructure would involve:

- Peer-reviewed and validated FPMs from literature, industrial practice, and academic repositories;
- Standardized representation of model inputs/outputs, equations, and assumptions;
- Domain-specific fine-tuning using chemical engineering knowledge graphs and expert annotations;
- Integration with PINNs and symbolic regression (SINDy) for equation discovery.

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

The potential impact is transformative, reducing FPM development from months to days and lowering barriers for smaller organizations. However, there will be also challenges such as data privacy in industrial FPMs, quality curation of training data, avoiding model plagiarism/overfitting to existing models, and human-in-the-loop validation. This is perhaps speculative at this point but grounded in current AI research trends, so it should be considered a feasible future direction worthy of research investment.

6. EDUCATIONAL AND INDUSTRIAL PERSPECTIVES

This section provides some educational and industrial perspectives, largely based on the academic and industrial experience of the authors (over 75 years combined).

6.1. Educational Perspective

Effectively training chemical engineers in FPM requires a curriculum that integrates foundational principles with computational techniques. As FPM often involves complex mathematical models, data-driven approaches and AI/ML methods, curricula should be designed to foster both theoretical understanding and practical skills.

Balancing theory and computation: Courses must continue to emphasize core principles (e.g., process synthesis, design and control, optimization, thermodynamics, transport phenomena, reaction engineering, unit operations) as the non-negotiable bedrock of the chemical engineering discipline. This must be coupled with rigorous training in computational skills, including numerical methods, programming, data analysis, and machine learning.[73] This dual focus ensures that engineers can develop models that are physically meaningful and also computationally efficient and effective.

Integration of AI/ML techniques: One could argue that AI and ML should not be taught as a separate, competing discipline. Instead, modules on neural networks, Gaussian processes, and reinforcement learning should be embedded within core process modeling and control courses.[11,73] Students should learn to critique the applicability, strengths, and limitations of various approaches, for instance, by using ML to create a surrogate model for a first-principles reactor simulation or to learn a residual term that corrects a simplified physical model.

Model conceptualization: Students find it hard to properly conceptualize a first-principles model based on an adequate textual problem description, even harder than coding and solving a model.[74] This challenge will be augmented by having to consider trade-offs between mechanistic and data-driven model parts and to find a balance. In hybrid modeling there are even more options to consider and compare, and to decide on the better one (s). The conceptualization of hybrid models is likely to remain a limiting factor in learning modeling skills by students, and it will require proper attention in teaching. In addition, it is very important to have a clear workflow in modeling (as shown in Figure 3).

Hands-on design projects: Project-based learning is of utmost important to ChE education.[75] Assignments should involve real (or simulated) industrial data, commercial process simulators (such as Aspen Plus, HYSYS, ChemCAD, gPROMS), and programming environments (e.g., Python, MATLAB). Such projects could involve developing a hybrid model for a (reactive) distillation column, optimizing a reactor using physics-informed models, or building a digital twin for a simple process, thus teaching the students various aspects of model calibration, validation, and deployment.

Using real industrial challenges for education purposes: Safety and optimization challenges of industry which combines FPM and DDM to overcome these challenges will help students to understand the value of these combined approaches in support of real life problem solving.

Critical thinking and model limitations: Education must go beyond equation-solving to emphasize the assumptions, boundaries, and economic trade-offs. Students must also learn to ask: *When is a data-driven approach sufficient versus when is a physics-based model necessary? What are the risks of extrapolation? What is the total cost of ownership of a model (over its lifecycle)?* This fosters adaptability and informed decision-making, preparing them for real industrial challenges, where the easy answer is rarely the best one.

Use of LLMs for modeling with critical assessment of the resulting models in educational programs. LLM could assist students in generating process models, equations, and simulation code, shifting the focus to analysis. As outputs often contain errors, this requires validation against first-principles, simulations, and experiments. Of course, curricula must enforce audits and ethical training to prevent over-reliance on LLM.

Physics-based education keeps an essential role in the LLM age. Paradoxically, as LLMs become more capable at generating mathematical models and scientific text, the need for chemical engineering education grounded in first-principles understanding becomes more critical, not less. Students and practicing engineers must develop:

- *Physical intuition* to understand why equations have particular forms, thus allowing engineers to evaluate LLM-generated models for reasonableness.
- *Domain expertise*, as knowledge of which assumptions are valid in different regimes enables critical assessment of LLM outputs.
- *Validation capabilities*, as the ability to design experiments and analyses allows proving or disproving LLM-generated hypotheses.
- *Trustworthiness assessment*, namely the judgment to determine when an LLM-generated model is validated sufficiently for deployment in safety-critical or economically important applications.

6.2. Industrial Perspectives

In the chemical process industry, the choice between various modeling paradigms is eventually driven by value, risk, and regulatory compliance, and not by academic trends. While AI tools are powerful assistants,[10] they should be carefully integrated to complement hard-won engineering expertise (not replace it). Preserving and leveraging key physical insights ensures safe, reliable, and optimized operations, while enabling understanding of process troubleshooting. Developing standards and best practices for hybrid modeling will facilitate broader adoption and acceptance.[58]

Industrial uses of models will always involve an assessment of the accuracy and relevance of the computed model responses. Models are used in an environment of decision making with multiple trade-offs. There are often several ways to achieve a particular goal. For instance, it may happen that the performance of a catalyst in a reactor does not entirely live up to its model-predicted activity. Then one may resort to making a (slight) change in operating conditions (e.g., temperature), as to force the catalyst to meet its expected performance. The value of a model goes beyond obtaining precise numerical values for the

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

model outputs. Getting an understanding of the uncertainties in model outputs and of the patterns of interaction between model outputs is just as important as knowing the absolute outcomes of model outputs. For mechanistic models there are two complementary roads to get information on interactions:

1 *Sensitivity analyses:* Derivatives of model outputs with respect to model parameters and inputs.
2 *Statistical analyses:* Obtaining the matrix of covariances of the parameter estimates.

When using locally a data-driven submodel within a mechanistic model, it is important that such a submodel is also offering the pertinent sensitivity and accuracy information.

First-principles models are routinely used for process design, scale-up, safety analysis, and troubleshooting.[14] However, contemporary sources confirm FPM's continued relevance in industrial practice (not due to inertia), but due to its irreplaceable role in safety requirements, regulatory compliance, and industrial innovation.[15−17] Their transparency is invaluable for diagnosing plant issues, gaining regulatory approval, and ensuring safe operations, especially during novel process upsets or start-ups where data is absent. Physical insight provides the guardrails within which data-driven models can safely operate.

Industrial adoption is not just about prediction accuracy on a test set; it involves a total cost of ownership (TCO) analysis. While data-driven models may have lower initial development costs for specific tasks (e.g., a soft sensor), first-principles models often prove more adaptable and cheaper to maintain over a process's long lifespan. Their inherent structure allows for easier adaptation to new feedstocks, operating conditions, or equipment changes without requiring a completely new data set and retraining—which is a significant advantage over pure black-box models.

The industrial software landscape is evolving to support this synergy.[59] Nowadays, the major process simulation vendors offer integrated hybrid modeling platforms (e.g., Aspen Hybrid Models), which allow embedding AI components within rigorous first-principles flowsheets.[63] This allows the development of "*digital twins*" (i.e., high-fidelity, validated models used for real-time optimization and operator training), which are almost invariably hybrid in nature, combining physics with real plant data.

The industry needs developed standards and best practices to allow a broader adoption of hybrid approaches, especially in safety-critical applications. This includes developing frameworks for validating data-driven components, quantifying uncertainty in hybrid model predictions, and ensuring that physical consistency is hard-coded and not just learned.[39,69]

7. CONCLUSIONS

This perspective argued that the foundational role of first-principles modeling (FPM) is not diminished, but is in fact more essential than ever in an era where AI and ML are rapidly expanding computational capabilities. FPM provides the critical attributes of physical correctness, interpretability, and reliable extrapolation, forming the trustworthy basis upon which powerful and safe hybrid models must be built. In contemporary practice, both approaches are essential and complementary. An integrated portfolio approach (using FPM for design and innovation, DDM for operations and control, and hybrid approaches where appropriate) reflects the current best practice in leading chemical companies. In all fairness, physics-informed modeling is not obsolete in the era of LLMs, rather it becomes essential for validating, understanding, and safely deploying models that may be generated with machine assistance. The real innovation lies not in AI replacing human expertise, but in human experts leveraging AI tools to extend their own capabilities.

The analysis presented herein moves beyond the simplistic FPM vs AI narrative. By deconstructing the model-building workflow, we have identified specific points where FPM knowledge is typically incomplete and where DDM can add maximum value (either through learning residual terms, approximating complex submodels, or creating physics-informed surrogates). This synergistic integration promises a new generation of models with enhanced predictive power, robustness, and safety, capable of driving forward the innovation in process design, process intensification, and real-time optimization. However, realizing this full potential requires continued effort and points toward a clear research agenda.

Addressing the following key open challenges will be crucial for the efficient widespread adoption of hybrid modeling as the new paradigm in chemical engineering:

- Enhanced education of students in proper conceptualization and understanding of (hybrid) models.
- Developing standardized validation and uncertainty quantification frameworks for the data-driven components of hybrid models.
- Creating robust methods for ensuring numerical stability and physical consistency (e.g., hard-coding conservation laws) within hybrid architectures.
- Establishing best practices for assessing the total cost of ownership over a hybrid model's full lifecycle.
- Building tools for the automatic or semiautomatic generation of hybrid models based on a high-level specification of the process and available data.

Ultimately, the rumors of the obsolescence of first-principles modeling have been greatly exaggerated. It is not even a question of replacement, but of intelligent collaboration and partnership. First-principles still lay down the foundational laws and offer deep insight, while data-driven approaches deliver adaptability and computational speed. Combining these strengths is the way forward to allow the creation of reliable, efficient, and genuinely innovative engineering solutions. Embracing this coexistence is clearly needed to effectively address the increasingly complex challenges facing modern industry and scientific progress.

AUTHOR INFORMATION

Corresponding Author

Anton A. Kiss − *Department of Chemical Engineering, Delft University of Technology, Delft 2629 HZ, The Netherlands;* orcid.org/0000-0001-5099-606X; Email: tonykiss@gmail.com, a.a.kiss@tudelft.nl

Author

Johan Grievink − *Department of Chemical Engineering, Delft University of Technology, Delft 2629 HZ, The Netherlands*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.iecr.5c04156

Notes

The authors declare no competing financial interest.

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

■ ACKNOWLEDGMENTS

The authors acknowledge the support of Farzad Mousazadeh (TU Delft) in reviewing the original paper and providing comments.

■ REFERENCES

(1) Marquardt, W. Trends in computer-aided process modeling. *Comput. Chem. Eng.* **1996**, *20* (6−7), 591−609.
(2) Kiss, A. A.; Grievink, J. Process Systems Engineering developments in Europe from an industrial and academic perspective. *Comput. Chem. Eng.* **2020**, *138*, 106823.
(3) Li, Q.; Somoza-Tornos, A.; Grievink, J.; Kiss, A. A. Challenges and opportunities for process intensification in Europe from a Process Systems Engineering perspective. *Front. Energy Res.* **2024**, *12*, 1340612.
(4) Bird, R. B.; Stewart, W. E.; Lightfoot, E. N. *Transport phenomena*; John Wiley & Sons, 1960.
(5) Bird, R. B.; Stewart, W. E.; Lightfoot, E. N. *Transport phenomena*, 2nd Ed., ed.; John Wiley & Sons, 2007.
(6) Aris, R. *Mathematical modelling techniques*; Pitman Publishing, 1978.
(7) Aris, R. *Mathematical modelling. A chemical engineer's perspective*; Academic Press, 1999, Vol. *1*.
(8) Hangos, K.; Cameron, I. *Process modelling and model analysis*; Academic Press, 2001, Vol. *4*.
(9) Dimian, A. C.; Bildea, C. S.; Kiss, A. A. *Integrated design and simulation of chemical processes* 2 nd Vol. *35*; Elsevier, 2014
(10) Chiang, L. H.; Braun, B.; Wang, Z.; Castillo, I. Towards artificial intelligence at scale in the chemical industry. *AIChE J.* **2022**, *68*, No. e17644.
(11) Bradley, W.; Kim, J.; Kilwei, Z.; Blakely, L.; Eydenberg, M.; Jalvin, J.; Laird, C.; Boukouvala, F. Perspectives on the integration between first-principles and data-driven modelling. *Comput. Chem. Eng.* **2022**, *166*, 107898.
(12) Qin, S. J. Survey on data-driven industrial process monitoring and diagnosis. *Annu Rev. Control.* **2012**, *36* (2), 220−234.
(13) Paris, R.; Beneddine, S.; Dandois, J. Robust flow control and optimal sensor placement using deep reinforcement learning. *J. Fluid Mech.* **2021**, *913*, A25.
(14) Pantelides, C. C.; Renfro, J. G. The online use of first-principles models in process operations: Review, current status and future needs. *Comput. Chem. Eng.* **2013**, *51*, 136−148.
(15) Pistikopoulos, E. N.; Tian, Y. Advanced modeling and optimization strategies for process synthesis. *Annu. Rev. Chem. Biomol. Eng.* **2024**, *15*, 81−103.
(16) National Academies of Sciences, Engineering & Medicine. Chemical Engineering *Challenges and Opportunities in the 21st Century*, 2022. https://www.nationalacademies.org/projects/DELS-BCST-18-01.
(17) Zhu, L.-T.; Chen, X.-Z.; Ouyang, B.; Yan, W.-C.; Lei, H.; Chen, Z.; Luo, Z. H. Review of machine learning for hydrodynamics, transport, and reactions in multiphase flows and reactors. *Ind. Eng. Chem. Res.* **2022**, *61*, 9901−9949.
(18) Vlachos, D. G. A review of multiscale analysis: Examples from systems biology, materials engineering, and other fluidsurface interacting systems. *Adv. Biochem. Eng.* **2005**, *30*, 1−61.
(19) Charpentier, J.-C. Among the trends for a modern chemical engineering, the third paradigm: The time and length multiscale approach as an efficient tool for process intensification and product design and engineering. *Chem. Eng. Res. Des.* **2010**, *88* (3), 248−254.
(20) Bard, Y. *Nonlinear parameter estimation*; Academic Press: New York, 1974.
(21) Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. *Global sensitivity analysis: the primer*; Wiley-Interscience, 2008.
(22) Rice, R. G.; Do, D. D. *Applied mathematics and modelling for chemical engineers*, 2 nd ed.; Wiley-AIChE, 2012.
(23) Rudd, D. F.; Watson, C. C. *Strategy of process engineering*; John Wiley & Sons, 1968.
(24) Biegler, L. T.; Grosmann, G.; E, I.; Westerberg, A. W. *Systematic methods of process design*; Prentice Hall, Upper Saddle River, 1997.
(25) Smith, R. *Chemical process design and integration*; John Wiley & Sons Ltd: Chichester, UK, 2005.
(26) Harmsen, J.; de Haan, A. B.; Swinkels, P. L. J. *Product and process design: driving sustainable innovation*, 2 nd ed.; De Gruyter, 2024.
(27) Zondervan, E.; Almeida-Rivera, C.; Camarda, K. V. *Product-driven process design. From molecule to enterprise*; De Gruyter, 2020.
(28) Edgar, T.; Himmelblau, D.; Lasdon, L. S. *Optimization of chemical processes*; McGraw-Hill Professional, 2001.
(29) Biegler, L. T. *Nonlinear programming: concepts, algorithms, and applications to chemical processes*; Philadelphia: SIAM, 2010.
(30) Seborg, D. E.; Edgar, T. F.; Mellichamp, D. A.; Doyle, F. J., III *Process dynamics and control*; Wiley, 2019.
(31) Kalman, R. E. Mathematical description of linear dynamic systems. *J. Soc. Ind. Appl. Math., Ser. A* **1963**, *1*, 152−192.
(32) Varma, A.; Morbidelli, M.; Wu, H. *Parametric sensitivity in chemical systems*; Cambridge University Press, 1999.
(33) Oberkampf, W. L.; Roy, C. J. *Verification and validation in scientific computing*; Cambridge University Press: Cambridge, 2010.
(34) Roache, P. J. *Verification and validation in computational science and engineering*; Hermosa Publishers, 1998.
(35) Schmal, J. P. *Dynamic chemical process modelling and validation: theory and application to industrial and literature case study*; Delft University of Technology, 2014.
(36) Benner, P.; Gugercin, S.; Willcox, K. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Rev.* **2015**, *57* (4), 483−531.
(37) Forrester, A. I. J.; Keane, A. J. Recent advances in surrogate-based optimization. *Prog. Aeronaut. Sci.* **2009**, *45* (1−3), 50−79.
(38) Dobbelaere, M. R.; Plehiers, P. P.; Van de Vijver, R.; Stevens, C. V.; Van Geem, K. M. Machine learning in chemical engineering: Strengths, weaknesses, opportunities, and threats. *Engineering* **2021**, *7* (9), 1201−1211.
(39) Samek, W.; Wiegand, T.; Müller, K.-R. Explainable Artificial Intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv* **2017**, *1*, 1−10.
(40) Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1* (5), 206−215.
(41) Gaddem, M. R.; Kim, J.; Matsunami, K.; Hayashi, Y.; Badr, S.; Sugiyama, H. Roles of mechanistic, data-driven, and hybrid modeling approaches for pharmaceutical process design and operation. *Curr. Opin. In Chem. Eng.* **2024**, *44*, 101019.
(42) Schweidtmann, A. M.; Zhang, D.; von Stosch, M. A review and perspective on hybrid modeling methodologies. *Digital Chem. Eng.* **2024**, *10*, 100136.
(43) Brun, R.; Kühni, M.; Siegrist, H.; Gujer, W.; Reichert, P. Practical identifiability of ASM2d Parameters - Systematic selection and tuning of parameter subsets. *Water Res.* **2002**, *36* (16), 4113−4127.
(44) Haseltine, E. L.; Rawlings, J. B. Critical evaluation of extended Kalman filtering and moving-horizon estimation. *Ind. Eng. Chem. Res.* **2005**, *44* (8), 2451−2460.
(45) Himmelblau, D. M. Applications of artificial neural networks in chemical engineering. *Korean J. Chem. Eng.* **2000**, *17*, 373−392.
(46) Venkatasubramanian, V. The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE J.* **2019**, *65* (2), 466−478.
(47) Beisheim, B.; Rahimi-Adli, K.; Krämer, S.; Engell, S. Energy performance analysis of continuous processes using surrogate models. *Energy* **2019**, *183*, 776−787.
(48) Bhosekar, A.; Ierapetritou, M. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Comput. Chem. Eng.* **2018**, *108*, 250−267.
(49) Daoutidis, P.; et al. Machine learning in process systems engineering: Challenges and opportunities. *Comput. Chem. Eng.* **2024**, *181*, 108523.
(50) Sutton, C.; McCallum, A. An introduction to conditional random fields. *Found. Trends Mach. Learn.* **2012**, *4* (4), 267−373.

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253

(51) Belkin, M.; Hsu, D.; Ma, S.; Mandal, S. Reconciling modern machine-learning practice and the classical bias−variance trade-off. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (32), 15849−15854.
(52) Fortuna, L.; Graziani, S.; Rizzo, A.; Xibilia, M. G. *Soft sensors for monitoring and control of industrial processes*; Springer: London, 2007. DOI: .
(53) Zhou, T.; Gani, R.; Sundmacher, K. Hybrid data-driven and mechanistic modeling approaches for multiscale material and process design. *Engineering* **2021**, 7 (9), 1231−1238.
(54) Saxena, A.; Celaya, J.; Saha, B.; Saha, S.; Goebe, K. Metrics for offline evaluation of prognostic performance. *IJPHM* **2021**, *1*, 4−23.
(55) Kiss, A. A.; Segovia-Hernández, J. G.; Bildea, C. S.; Miranda-Galindo, E. Y.; Hernández, S. Reactive DWC leading the way to FAME and fortune. *Fuel* **2012**, *95*, 352−359.
(56) Elsheikh, M.; Ortmanns, Y.; Hecht, F.; Roßmann, V.; Krämer, S.; Engell, S. An approach to dependable hybrid modeling with adaptation of the domain of validity. *Comput.-Aided Chem. Eng.* **2023**, *52*, 1753−1758.
(57) Zuo, R.; Li, Y.; Wei, S.; Li, Z.; Shi, T.; Du, W.; Shen, W. Calibration-augmented and mechanism-driven deep learning hybrid framework for modeling actual distillation processes. *Ind. Eng. Chem. Res.* **2025**, *64*, 3856−3870.
(58) Gargalo, C. L.; Malanca, A. A.; Aouichaoui, A. R. N.; Huusom, J. K.; Gernaey, K. V. Navigating industry 4.0 and 5.0: The role of hybrid modelling in (bio)chemical engineering's digital transition. *Front. Chem. Eng.* **2024**, *6*, 1494244.
(59) Cai, T.; Fang, J.; Daida, S.; Lou, H. H. Review of synergy between machine learning and first principles models for asset integrity management. *Front. Chem. Eng.* **2023**, *5*, 1138283.
(60) Chakraborty, A.; Serneels, S.; Claussen, H.; Venkatasubramanian, V. Hybrid AI models in chemical engineering − A purpose-driven perspective. *Comput.-Aided Chem. Eng.* **2022**, *51*, 1507−1512.
(61) Rajulapati, L.; Chinta, S.; Shyamala, B.; Rengaswamy, R. Integration of machine learning and first principles models. *AIChE J.* **2022**, *68*, No. e17715.
(62) Shah, P.; Pahari, S.; Bhavsar, R.; Kwon, J. S.-I. Hybrid modeling of first-principles and machine learning: A step-by-step tutorial review for practical implementation. *Comput. Chem. Eng.* **2025**, *194*, 108926.
(63) Aspen Technology, *Aspen Hybrid Models*, 2023. www.aspentech.com/en/products/engineering/aspen-hybrid-models (accessed 23 Auguest 2025).
(64) Khor, C. S.; Elkamel, A.; Shah, N. Optimization methods for petroleum fields development and production systems: A review. *Optim Eng.* **2017**, *18* (4), 907−941.
(65) Raissi, M.; Perdikaris, P.; Karniadakis, G. E. 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.
(66) Karia, T.; Lastrucci, G.; Schweidtmann, A. M. Kolmogorov Arnold Networks (KANs) as surrogate models for global process optimization. *Syst. Control Trans.* **2025**, *4*, 1371−1376.
(67) Raissi, M.; Karniadakis, G. E. Hidden physics models: Machine learning of nonlinear partial differential equations. *J. Comput. Phys.* **2018**, *357*, 125−141.
(68) Thompson, M. L.; Kramer, M. A. Modeling chemical processes using prior knowledge and neural networks. *AIChE J.* **1994**, *40* (8), 1328−1340.
(69) von Stosch, M.; Oliveira, R.; Peres, J.; de Azevedo, S. F. Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Comput. Chem. Eng.* **2014**, *60*, 86−101.
(70) Wilson, Z. T.; Sahinidis, N. V. Automated learning of chemical reaction networks. *Comput. Chem. Eng.* **2019**, *127*, 88−98.
(71) Sharma, N.; Liu, Y. A. A hybrid science-guided machine learning approach for modeling chemical processes: A review. *AIChE J.* **2022**, *68*, No. e17609.
(72) Jia, Z.; Ierapetritou, M. G. Modeling and optimization of refinery operations considering uncertainty. In *Process Systems Engineering*, Pistikopoulos, E. N.; Georgiadis, M. C.; Dua, V., Eds.; Wiley, 2010. .
(73) Wu, M.; et al. An artificial intelligence course for chemical engineers. *Educ. Chem. Eng.* **2023**, *45*, 141−150.
(74) Somoza-Tornos, A.; Franke, M.; Haringa, C.; Kiss, A. A.; Mousazadeh, F.; Özkan, L.; ten Kate, A.; van Ommen, J. R.; Zondervan, E.; Grievink, J. Challenges for modelling in chemical engineering education in the Netherlands. In *ESCAPE 35 - European Symposium On Computer Aided Process Engineering, Book Of Short Papers, 273−275, EUROSIS-ETI/KU Leuven/BioTeC+*, Van Impe, J. F. M.; Léonard, G.; Bhonsale, S. S.; Polańska, M. E.; Logist, F. F.; KU Leuven, 2025.
(75) Kiss, A. A.; Webb, C. The Manchester perspective on using the Design Project to enhance the education of chemical engineering students. *J. Chem. Technol. Biotechnol.* **2021**, *96*, 1453−1464.

https://doi.org/10.1021/acs.iecr.5c04156
Ind. Eng. Chem. Res. 2026, 65, 4234−4253