
ARTIFICIAL INTELLIGENCE IN CHEMICAL ENGINEERING

WHITE PAPER



**European Federation of
Chemical Engineering**

**Société Française de
Génie des Procédés**



FOREWORD



Artificial Intelligence and Machine Learning have had in the recent past a tremendous impact on chemical engineering. On the long run they can radically change the way in which we perform research on chemical processes, we design, scale-up and optimize chemical reactors and pieces of equipment and they can potentially introduce new paradigms on chemical process control and operation.

Many are the challenges ahead for our community that can be addressed by the European Federation of Chemical Engineering (EFCE) and its Working Parties and Sections. These can be a successful journey only if the European Chemical Engineering community, both academic and industrial, is fully engaged.

The European Forum on New Technologies is ideal event to discuss such challenges as it offers a one-day event dedicated to new scientific developments where industrialists and academics can exchange views related to a specific chemical engineering challenge.

That is why we have brought together a number of visionary speakers at this event to present the potential and challenges ahead. There was the opportunity to discuss the fundamentals of artificial intelligence, deep learning and machine learning, as well as their applications to process modelling, control and the building of digital twins.

The 5th European Forum on New Technologies was organized in Paris on 13 December 2024, gathering experts, professionals, and researchers from academia and industry (with about a 50/50 representation academia/industry and no less than 42 companies taking part) to discuss the role of Artificial Intelligence in Chemical Engineering.

The large attendance (130 participants) coming from 17 countries confirms that Artificial Intelligence (AI) is a hot topic that concerns many areas of chemical engineering.

The program which was structured in three sessions followed by a round-table held at the end of the day covering different aspects of AI in Chemical Engineering.

This White Paper is based on the oral presentations given during the day and aims to summarize the discussions that took place. It describes the applications of AI in different areas of chemical engineering and provides guidelines and warnings about the precautions that should be taken when implementing it.

Martine POUX
Coordinator of the EFNT
EFCE Scientific Secretariat
SFGP General Delegate



STARTWON



02

Foreword

03

Contents

04

Overview of artificial intelligence for process engineering

08

Generative artificial intelligence and transformation in process engineering

11

Artificial intelligence applications in chemical engineering

16

Round Table

- Can AI enhance creativity in process engineering?
- How can AI be better integrated into chemical engineering curricula?
- What is the future of AI-human collaboration?

22

Acknowledgments

23

Appendix

OVERVIEW OF ARTIFICIAL INTELLIGENCE FOR PROCESS ENGINEERING

ARTIFICIAL INTELLIGENCE & MACHINE LEARNING IN THE PROCESS INDUSTRY: WHERE WE ARE AND WHERE WE ARE GOING

With AI, complex, non-linear objective function tasks, such as identifying the effect of temperature, flow or pressure on yield in a distillation column, that previously stumped chemical engineering graduates, can now be effectively addressed, with machines leveraging sensor data and employing feature engineering to identify critical factors, supported by explainable AI techniques such as parallel coordinate and Shapley plots. GenAI tools can now easily apply appropriate algorithms and predictive analytics to deliver strong results, making the democratization of advanced analytics in process engineering a reality.

Challenges remain, however, in applying these innovations within the industrial sector, notably due to weight of technical debt, rigid automation hierarchies as companies opt to retain foundational systems and reduce costs, and conservative practices, particularly in the chemical sector, driven by security concerns. Sensor reliability and the slow pace of sensor innovation further constrain progress. While significant opportunity resides in valuable data sources such as PIDs, process flow diagrams (PFDs), and flow-sheets, their integration with AI remains limited, and data-preparation for AI use also poses significant hurdles. Industry demand focuses on practical applications such as patent consolidation, literature-driven technology generation, and cost optimization, contrasting with academia's emphasis on complex models. Looking ahead, advances such as increased integration and AI-assisted troubleshooting, automated documentation, digital twins, VR environments, and self-optimizing systems, are expected to transform operations within the next two to five years, although foundational upgrades remain a current bottleneck.

DIFFERENT WAYS IN WHICH ARTIFICIAL INTELLIGENCE CAN BE USED WITHIN THE ORGANIZATION TO OPTIMIZE INDUSTRIAL PROCESSES

Industrial firms often underutilise their data, with access largely limited to specialists rather than operational teams. Unlocking value requires broad data utilisation alongside advanced analytics, including AI, statistics, and machine learning. Key barriers here include fragmented data systems, traditional data structuring, and the challenge of acquiring near real-time, contextualised data for meaningful decision making.

Effective application of AI hinges on establishing a strong business context, combining diverse data types such as time series from sensors and traceability data from production campaigns, and bridging the expertise gap between data specialists and process operators. Integrating heterogeneous data remains challenging due to differing formats, varying levels of expertise, and a lack of data proficiency among operational staff, particularly in resource-constrained environments. The deployment of AI in process manufacturing requires careful consideration of both operational technology and cloud-based systems, and embedding AI directly into applications allows users to benefit without deep technical knowledge, which streamlines operations. In complex industries, such as speciality chemicals and dairy, tracking genealogy from raw materials to final product is vital, as early-stage decisions impact end-quality.

Proper data structuring at the right granularity underpins robust model development and effective processing and aggregation of data, often through combining process expertise with machine learning. Model optimisation and lifecycle management, encompassing deployment, monitoring, and retraining with updated data, is essential for sustained value. Reducing model parameter redundancy, through consolidating correlated sensor data and focusing on key control parameters like vacuums, further enhances model efficiency, enabling more precise variability analysis and optimized process control tailored to manufacturing environments.

ARTIFICIAL INTELLIGENCE TOOLS FOR PROCESS MODELLING AND SIMULATION: A CRITICAL OVERVIEW ON POTENTIALITIES AND LIMITATIONS SO FAR

AI is drawing increasing attention in Chemical engineering, driven by broad interest in its potential benefits, academia's aim to solve related overarching challenges, and the industry's push for automation and improved efficiency.

A key challenge in process engineering has long been optimizing complex systems, particularly through Mixed-Integer Non-linear Programming (MINLP). While past hurdles like scheduling and control have by now been largely resolved, today's focus lies on designing plants capable of handling uncertainty, especially given variable renewables and decentralised, modular manufacturing. Black-box approaches help to integrate diverse components, addressing biological complexity and uncertainty often missed in practice, while academia emphasizes analytical models and industry leans on empirical, case-specific solutions.

Rather than questioning the reliability of data-driven methods, critical examination reveals that it is more productive to compare them with alternative approaches. Available tools, including emulation platforms, ProSim, and Aspen, are available for use in the field, which require thoughtful application of sampling strategies, modelling techniques, and crucially, data pre-treatment to ensure reliable outcomes.

A significant barrier lies in "trust stagnation", which undermines confidence in simulations, despite the fact that many chemical processes are less non-linear than once believed. Simplistic approaches may not suffice when assessing chemical plant output, and surrogate models must, in these cases, be cautiously implemented, as errors in one part can cascade through interconnected systems, distorting mass and energy balances. In dynamic models, system response time also represents a critical constraint, particularly when solely concerned with data-driven approaches. Ultimately, recognizing system-specific limitations and potentials is here essential, as conclusions on feasibility may differ across contexts.

DISCUSSION

While GenAI holds promise for advancing industrial processes, several challenges temper expectations for its near-term deployment. A central concern is the lack of explainability in current AI systems, which limits trust and raises questions concerning how much responsibility can be delegated to such technologies. Integrating documentation and contextual information into GenAI and machine learning systems is seen as essential to improving situational awareness and reliability.

Experts agree that fully autonomous AI-driven process management remains a distant prospect, with safety, ethical, and regulatory considerations to be addressed prior to real-world implementation, particularly in sensitive environments such as industrial plants. Current AI architecture, which includes large language models, is not designed to operate automated facilities, and significant technological hurdles remain. Leading research, such as that presented at recent AI conferences, highlights both the limitations of existing systems and the pathways to potential solutions. Overall, GenAI's role in automation is expected to evolve gradually, and any advances will require continuous reassessment as the technology matures.

The effective implementation of artificial intelligence in industrial and laboratory settings hinges on the availability of appropriate and sufficient data. While process manufacturing industries such as chemical, oil, and gas benefit from long-established systems for data collection, the agro-industry and food and beverage sectors often face more fragmented scenarios, particularly in smaller facilities with autonomous but disconnected control systems. Although advances in technology are improving access to data, key considerations around the operational technology layer remain critical to ensure accuracy.

A common challenge here lies in defining “right data”. Industrial operations typically focus on process data, including temperature, pressure, and time, derived through control systems.

However, this focus typically overlooks equally important product-related information, often obtained from laboratory analysis. Understanding the interaction between product and process is crucial, as maintaining reliable mass balance, a fundamental metric in this field, is essential for accurate calculation of yield and performance optimization with respect to control, laboratory, maintenance, and other related data sources. To achieve greater accuracy, real-time product data is increasingly prioritised over commonly delayed laboratory experiment result data. Work in this field is ongoing, with experts implementing online analysis tools that utilise near-infrared chromatography, which is crucial for robust monitoring and control precision.

Beyond data collection, maintaining general data integrity is of critical importance. Even when sensors appear to be properly functioning, subtle deviations may undermine data reliability, a phenomenon which underscores the importance of combining engineering expertise with statistical methods, data reconciliation, and other verification techniques to validate and contextualise data. Ultimately, successful AI deployment will be contingent not only upon data quantity but also quality, as well as integration capabilities across multiple systems.

A fundamental distinction between physical and data-driven models lies in their ability to provide explainability and establish causality, an aspect which becomes critical when operators rely on models to inform decision-making. While physical models offer clear, causally grounded insights, data-driven models often fall short in transparency, which can limit user trust and acceptance.

Ensuring model explicability has hence become a central focus in laboratory research, particularly in the context of AI’s turbulent history. So-called “AI winters” of the past were marked by a stagnation of research efforts, and these events now serve as cautionary reminders of risks involved when confidence in AI tools erodes. Maintaining explicability is hence deemed essential to prevent repeats of this cycle.

Techniques such as the Shapley approach have emerged as promising tools to improve explainability, although they fail to address causality. Beyond technical solutions, fostering trust also requires the careful definition of a model’s operational boundaries. Users must understand where models perform reliably and where caution is warranted. Providing clear guidance here on appropriate input and application domains helps avoid misuse and promotes effective usage.

In summary, designing models that balance effective control, user interaction, and transparency remains a significant challenge, but is one that must be addressed to ensure responsible and sustainable AI integration into industrial processes.

In the context of industrial processes, whether in reactors, distillation units, or other equipment, the choice of modelling approach plays a pivotal role in achieving meaningful insights. Machine learning and statistical models are often viewed as distinct from physical principle-based models, yet, when thoughtfully combined, they can prove highly complementary. On one end of the spectrum here lies the “black-box” approach, while the phenomenological approach rests on the other. Between these poles exists a rich space of alternative methodologies that blend foundational principles to enhance precision and reliability.

A critical insight to be gleaned here is that much of the value of AI modelling does not stem solely from selecting or training the model itself, but from feature engineering. Here, domain expertise in physics and chemistry takes on essential importance, and can be integrated through strong feature engineering informed by robust understanding of processes involved. Hence, effective pre-processing serves as the bridge between physical insight and machine learning.

A critical concern in applying AI to industrial processes is the risk of introducing human biases, particularly during feature selection or data interpretation stages. This raises both technical and philosophical questions. It is essential to understand the limitations of the methodologies in use and to approach machine learning, statistical models, or data-driven models, from a position of sufficient understanding of the tools involved. Without a solid grasp of related functions, bias could easily be introduced into results. While feature engineering plays a major role here, the influence of bias can at times be diminished by selecting appropriate algorithms to help balance discrepancies. Nonetheless, defining clear performance criteria remains the responsibility of those developing the models. Even if some bias enters the model through feature engineering, reaching targeted performance outcomes suggests the meeting of objectives, although this does not automatically guarantee reliability.

Another crucial factor here is data quality. New or more sophisticated models are not inherently more trustworthy, and underlying data quality remains of critical importance. Industrial settings also present their own challenges, including engaging with operators, on which the importance of robust methodologies can not be overstated. Without dependable methods, AI is at risk of being built upon flawed foundations, and best practices such as using models to identify sensor drift or to perform mass-balance calculations represent essential safeguards to ensure data integrity and trustworthy AI applications.

GENERATIVE ARTIFICIAL INTELLIGENCE AND TRANSFORMATION IN PROCESS ENGINEERING

○ TRANSFORMING PROCESS ENGINEERING WITH GENERATIVE ARTIFICIAL INTELLIGENCE

GenAI is already making notable contributions to the computing industry, particularly in tasks such as automated error correction. In process engineering, this capability holds obvious promise, particularly for managing complex engineering diagrams such as P&IDs that can span thousands of pages. Beyond simply catching mistakes, GenAI has the potential to improve maintainability, controllability, and sustainability. It could, for example, assist engineers by suggesting how to specify, operate, and design target systems based on available data. GenAI's role could also extend to hazard identification, helping to generate preliminary safety assessments and saving valuable time.

While tools such as ChatGPT can already generate process suggestions, their effectiveness in creating practical engineering solutions remains limited. Progress in this area will require close collaboration between researchers and industry to develop GenAI tools that are both reliable and domain specific. Key components for success here include access to substantial data, effective information representation, and the use of appropriate model architectures, particularly transformer models, alongside integration of domain knowledge from physics and material science. The established roadmap for these efforts aims to address four key challenges, including auto-completion of flow-sheets, PFD-to-PID translation, auto-correction of PIDs, and AI-augmented HAZOP.

A central question is whether chemical and process engineering qualifies as a “big data” domain, as current literature suggests that data availability is limited in the field. As an interim solution, hybrid models that combine physical principles with machine learning have emerged, which reduce data requirements while enhancing generalization.

The engineering community generates vast amounts of data, from flow-sheet simulations and property data to academic research outputs. The challenge here lies in the heterogeneous nature of this data, which is often difficult to access, not readable by machines, or restricted by confidentiality. To overcome this, collaboration with industry and dedicated data-gathering initiatives, such as Digico, a digitization companion, are essential. Once data is acquired, it must be prepared for machine learning models. Two main representations dominate here, including graph representation, where equipment items are treated as nodes and their connections as edges, and string representation, which is favoured by many GenAI models for its simplicity. Selecting the appropriate machine learning architecture is equally important. Classical transformers are well suited for sequential text data, while graph neural networks excel in graph-based data contexts. The choice here should align with the type of data being used.

Although rule-based approaches offer transparency and efficiency, they face challenges: rule implementation demands deep expertise, and even minor errors can lead to system failures or overlooked edge cases. Despite these complexities, certain rules can be easily implemented and may apply broadly across multiple PIDs to improve safety. Machine learning offers the added promise of automatically deriving rules from existing PIDs, which could significantly streamline design and testing. The broader success of models such as ChatGPT lies in their ability to integrate across diverse tasks and languages, and engineers require such tools to integrate multi-modal information and provide interconnected process insights to better generate and refine ideas.

Looking ahead, research aims to establish a PILOT process intelligence and optimization tool, a multi-agent system where large language models can interact and collaborate on specialised tasks, interfacing with databases and simulation tools to drive advances in process synthesis.

In summary, data access is of critical importance in the field. As most university researchers do not possess their own sizeable engineering datasets, partnerships with industry players are essential. Without prioritizing access to digital engineering data, the field runs the risk of falling behind in the competitive landscape of innovation, and of missing out on key opportunities for business while transforming engineering practice for the better through the incorporation of GenAI into its methodologies.

○ PROCESS FLOW-SHEET GENERATION BY ARTIFICIAL INTELLIGENCE: MOTIVATION & CURRENT STATE

EDF focuses not on the sale of individual processes, but on delivering electric power systems: a mission that spans weather conditioning, wastewater treatment, and energy conversion. While nuclear reactions share roots with chemical processes, they bring distinct challenges, such as mass imbalances, which necessitate cautious selection and process design.

To improve process design, EDF has long relied on computer-aided engineering, enabling engineers to select and simulate various processes tailored to specific applications. Today, the approach here is evolving from one of a traditional iterative design cycle to one of generative paradigm, where engineers define desired performance outcomes, while AI assists in determining optimal configurations. Compared to traditional methods, the benefit of this approach may be illustrated through a recent case which explored the CO₂ Brayton cycle as a potential alternative to conventional steam cycles. Using simulation tools, engineers here conducted parametric studies across temperatures and pressures to successfully optimize net cycle efficiency.

To facilitate AI-driven process design, the laboratory developed a specialized, compact language that represents PFDs as character sequences. Training models to work with this language requires them to grasp its syntax, as random letter combinations do not constitute valid words or processes. A recurrent neural network was employed by researchers to this end, capable of generating process sequences by predicting next equipment items, such as turbines, based on prior elements. While the model was trained without explicit guidance, specific filtering rules ensured that the resulting process diagrams were coherent, interconnected, and meaningful. Fine-tuning and optimisation marked the next phase of research, in which the AI-generated PFDs were simulated and optimised using in-house simulation, translating the AI's terminology into the simulator's language and setting necessary optimisation parameters. Research here remains ongoing, and going forward, robust communication standards and data formats could enable the integration of established process simulators like ASPEN into this new and innovative AI-driven workflow.

To date, two key objectives have been achieved here: power cycle efficiency and shaft power, which are both critical for maximizing output. The algorithm's ability to generate a diverse array of processes highlights its potential to produce novel designs, and comparisons with expert designs indicate promising alignment, underscoring the system's capacity to incorporate established heuristics and provide innovative and unconventional solutions.

Looking ahead, EDF aims to systematize the computational generation and evaluation of numerous processes, which will add a new layer of automation to process design. The goal here is not to replace engineers, but to free them to focus on more critical tasks, such as defining and analysing problems, as well as formulating and evaluating solutions, with these areas remaining firmly in human hands, beyond the capabilities of machines. However, two critical challenges remain to be addressed. The first of these is data availability, with molecular design benefitting from abundant and well-documented datasets, while in contrast, process engineering lacks comparably rich datasets, partly due to how processes are recorded, as well as the nature of their specific properties. The second of these is process evaluation. While generating possible process flows through AI is becoming increasingly feasible, assessing their performance remains a considerable challenge, while optimization can be highly resource-intensive, making the development of more efficient approaches, such as benchmarking frameworks, a high priority task for current research.

DISCUSSION

Non-converged solutions are common in process simulation, especially when tools limit access to the underlying code. To address this, hard-coded simulations such as Hinao's have been used, although they still produce some failures. Some non-convergence here stems from attempting to simulate invalid processes, which can simply be excluded during design and generation. For issues of purely convergent nature, the perspective has shifted: instead of discarding all failed solutions, partially successful outcomes, such as those indicating 90% efficiency metrics, are now seen as valuable, and the introduction of penalty functions represents a potential strategy to handle such cases.

The application of autocompletion features in PNI offers promising insights but presents significant challenges due to the variability of control structures across industrial plants. While models can be completed to reflect specific scenarios, they may not generalize accurately, as control objectives may differ greatly. For example, pressure control in a flash unit may conflict with maximizing flow if the stream returns to a reactor, and in heat exchanges or condensers, cascade control strategies may be required to reconcile temperature and flow objectives. Such complexities underscore the importance of considering system-specific logic when implementing PNI autocompletion. Experts have expressed confidence that initial proofs of concept will, when built upon by a collaborative research community, result in the discovery of new, innovative, and effective solutions.

Recent research has explored the use of superstructure optimization methods to improve model outcomes. Comparative evaluations of three optimization approaches have indicated that the superstructure method currently delivers the most promising results, outperforming alternatives such as evolutionary programming. Although further improvements in training and simulation times may enhance the superstructure's performance, current solutions already approximate near-optimal configurations.

Looking ahead, the integration of AI with established domain knowledge is seen as a critical research frontier. While rule-based systems alone are restricted by limitations, the process industry anticipates the development of hybrid systems that blend AI capabilities with expert oversight, with the incorporation of robust data management through "data room" concepts is expected to play a central role in enabling these innovations.

The parallels between P&IDs and electronic circuits present another compelling opportunity for the application of GenAI. Significant progress has been made in electronic circuit design using generative models, supported in part by the availability of extensive open-access databases. By contrast, similar resources are notably scarce in the domain of P&IDs, limiting the ability here to train and validate AI models at scale.

The proprietary nature of industrial process data poses a key challenge in this respect. Many P&IDs, particularly those associated with sensitive sectors such as nuclear energy, are tightly guarded and unlikely to be released publicly, creating a critical bottleneck for developing and generalizing AI applications in process industries. Additionally, the structural and functional differences between PFDs and P&IDs must be acknowledged. While process engineers typically develop one form, project engineers are responsible for the other, reflecting distinct scopes and purposes. In large projects, approximately 20% of P&IDs are developed by engineering firms, while the remaining 80% often originate from various technology licensors. Consequently, there is limited scope for intervention in this area.

Despite these challenges, collaborations with engineering firms have enabled progress by focusing on simpler, modular tasks, with standardization and improved data accessibility representing critical factors towards successful AI integration in the field.

ARTIFICIAL INTELLIGENCE APPLICATIONS IN CHEMICAL ENGINEERING

● HYBRID MODELLING AND DIGITAL TWIN IN CLINKER PRODUCTION

Industrial integrators such as Technord employ hybrid modelling approaches in chemical engineering, integrating data-driven techniques with physical mass and energy balance to enhance process optimisation. This approach merges statistical analysis, regression, and traditional methods with differential equations, creating a more accurate solution space compared to purely data-driven modelling. Focus here lies on developing software sensors to support decision-making in industrial operations, emphasizing intelligent maintenance and pattern recognition while assessing the impact of machine quality on process development by bridging data silos across systems.

A current project in clinker production uses advanced process control for predictive KPIs and early detection, aiming for optimized energy use and predictive maintenance. Three key pillars drive this initiative: defining the use case, ensuring the availability of high-quality data, and securing support for change management. The goal is to optimize the combustion process in cement production, particularly by efficiently storing limestone and lime to minimize energy consumption in the kiln, reducing operational risks such as clogging and excessive energy input.

Data analysis plays a critical role in monitoring and adjusting energy usage in this field. By focusing on minimizing fluctuations around target CO₂ levels, researchers aim to prevent underutilisation or overloading of energy systems, which can disrupt production, as well as reducing costs. After months of monitoring work, involving crucial high-quality data modelling techniques revolving around KPIs concerning quality in real time, the predictive model employed achieved an 84% accuracy rate, with plans since being established to implement a closed-loop system supported by structured training for operators, with a focus on ensuring that process engineers recognize that the introduction of AI serves as a supportive tool rather than a threat to their employment. This system leverages AI to enhance performance metrics, optimize operational processes, improve energy efficiency, and stabilize product quality, while adapting to changing market requirements. Engaging in such data-driven projects necessitates critical evaluation of information to guide decision-making, and in the incident production sector, which operates on slim profit margins, the importance of volume and the need for increased optimization to maintain competitiveness is increasing.

● USING MACHINE LEARNING FOR ONLINE 3D CHARACTERIZATION OF CRYSTALS IN SUSPENSION

Crystal morphology plays a critical role in material processing. For instance, two powders of identical polymer type and mass but differing crystal shapes may exhibit vastly different behaviours: equant crystals flow smoothly and compactly, while needle-like crystals aggregate into clumps and occupy greater volumes, despite featuring identical mass and density. This reality underscores the need for tailored characterization approaches across diverse morphologies. Crystals can be broadly classified into three morphological categories for characterisation: spherical approximations, with cubic crystals such as sodium chloride effectively described through a single dimension, this being diameter; cylindrical approximations, with needle-like crystals requiring two-dimensional length and width characterization; and cuboidal approximations, with plate-like particles demanding a full three-dimensional characterisation, incorporating length, width, and thickness.

To date, only two academic instruments are capable of performing true three-dimensional characterization of plate-like crystals: a recently developed offline device at the University of Manchester, and an online instrument developed in-house by a Swiss laboratory. This device, currently seeing research usage, is known as the Dual Imaging System for Crystallisation Observation (DISCO). It operates by circulating suspension from a reactor through a glass cell, where two perpendicular cameras, illuminated for enhanced contrast, capture particle images which allow for effective particle analysis and subsequent three-dimensional reconstruction based on matched image contours. However, caveats are present in this case, with the measuring of platelet length raising challenges, and with the same particles potentially appearing differently across images based on their orientations at the moment of capture.

Single-camera systems can estimate only up to two dimensions, often with orientation-dependent variability. The addition of a second camera in this case improves reconstruction accuracy, but may on occasion overestimate thickness when combining projections. Understanding these effects is crucial towards final accuracy in the measurement of different populations and in employing various length estimation algorithms. Recognising these limitations, research teams have developed a comprehensive test set covering a wide range of particle types and morphologies, which enables systematic evaluation of length estimation algorithms under certain constraints that limit the range of particle sizes.

Experimental workflows here involve repeated sampling of the same spatial region, the inclusion of non-cuboidal particles, and simulated measurements. From the resulting contours and 3D reconstructions, which encompass circularity, convexity, and area and volume scale, both 2D and 3D pre-processing steps can employ intercorrelation filtering, recursive feature elimination, and hyper-parameter optimisation, followed by a seven-fold cross-validation and ensemble averaging process across 10 models. The culmination of this work was a neural network model in which the average error metric was reduced by 33% when compared to conventional approaches. Additionally, research here has probed the relationship between crystal morphology and model performance, and has demonstrated superior precision of the machine learning approach in characterizing plate-like particles.

However, the computational model does not fully replicate real-world conditions, omitting factors such as blur and assuming uniformly distributed particle orientations. To bridge this gap, experimental validation is being pursued using an innovative analytical standard fabricated via photolithography techniques, which enable precise and previously unavailable ground-truth measurements.

PREDICTING SOLUBILITY LIMITS WITH MACHINE LEARNING

Regarding property prediction models, researchers emphasize the relationship between molecular structure and properties. Accurate molecular property predictions are fundamental to designing efficient process flow-sheets and unit operations, particularly when dealing with novel molecules. Research here also emphasises the prediction of solubility limits, specifically within the pharmaceutical industry, where understanding the solubility of active pharmaceutical ingredients in various organic solvents is vital in the drug design process.

Two key properties lie at the heart of current analysis in this field: solvation-free energy and solid solubility. Both have profound implications for optimizing processes such as organic synthesis, liquid-liquid extraction, and continuous processing, as they directly influence the efficiency and feasibility of pharmaceutical manufacturing.

Central to the development of predictive models is the challenge of effectively representing molecular structures within a machine learning framework. Graph neural networks have emerged as a powerful solution here, and are particularly suited for molecular systems. In this approach, atoms serve as nodes and bonds as edges, each characterized by specific atomic features, including atomic number and mass, as well as bond features such as bond type. Graph neural networks facilitate machine learning by facilitating the exchange of information between neighbouring atoms and bonds, ultimately generating latent representations that correlate with target molecular properties.

For solvation-related property analysis, researchers consider it essential to consider both the solute and solvent as separate entities within the neural network. This separation is accomplished using two distinct message-passing neural networks, one for the solute and another for the solvent, whose outputs are then concatenated to predict the desired property. Extending this framework to handle solvent mixtures introduces additional complexity, and to ensure that the neural network provides consistent results regardless of the order of solvents, the network incorporates solvent embeddings weighted by molar fractions which are summed to enable comprehensive embedding and to guarantee permutation invariance for solvent mixtures, allowing for effective prediction of solubility properties of solutes within target mixtures.

While there exists an abundance of data concerning certain molecular properties, one major challenge in this field is the limited nature of P&ID data specifically, particularly for solubility properties in organic solvents. Hence, urgent need is observed by research for innovative approaches such as transfer learning and hybrid modelling. While substantial datasets exist for solvation-free energies derived from vapour-liquid equilibria and activity correlations, this data is not extensively applicable to pharmaceutical contexts. Hence, research objectives now focus on the development of a robust model that leverages data from common solvent molecules to enhance pharmaceutical applications.

Researchers have here adopted transfer learning strategies, combining large, lower-accuracy quantum chemistry datasets with smaller, higher-accuracy experimental measurement datasets. Initially, the neural network is trained on data derived from quantum chemistry, and is then fine-tuned using experimental data to significantly improve predictive accuracy, particularly for solvation-free energies.

For solid solubility prediction, where quantum chemistry methods fall short, hybrid models have been established that combine machine learning with established thermodynamic principles to accurately predict the solid solubility. A striking application of these predictive models is in the automated design of dyes with tailored properties. Utilizing a robotic laboratory platform, researchers have employed AI to autonomously generate novel dye structures with enhanced attributes. This system integrated property prediction models to assess candidate molecules, devise synthesis pathways, and conduct synthesis and experimental validation. Over multiple iterations, the platform successfully developed dyes with improved features, such as enhanced absorption and optimized water partitioning.

These machine-learning driven property prediction models have become indispensable tools in modern molecular design, enabling researchers to navigate vast chemical spaces with superior efficiency. Without such predictive capabilities, many newly generated molecular structures would lack practical value. Ultimately, such trends underscore the essential role of AI in the acceleration of discovery and general innovation in the field.

DEEP-LEARNING METHODS FOR THE IMAGE-BASED ASSESSMENT OF THE PHYSICAL STABILITY OF FORMULATED LIQUIDS OF INDUSTRIAL INTEREST

Formulated liquids of industrial significance are complex, consisting of intricate mixtures of numerous components, including water, surfactants, perfumes, thickeners, dyes, and, crucially, stabilizers, which represent the focus of current research, as ensuring the stability of these products is essential for product performance and acceptance. The four principal types of instability typically encountered here include phase separation, where immiscible phases separate, producing distinct layers; creaming, where less dense phases rise upward under gravitational pressure; cracking, which manifests as visible fissures within the product; and flocculation, the aggregation of dispersed particles into clumps or flocs.

Despite the critical nature of stability assessment, no universal method exists for assessing the stability of such industrial liquids. Conventional approaches include nuclear magnetic resonance microscopy, automated turbidity measurements, and optical techniques. However, these methods face several limitations, including high equipment costs and the need for specialized personnel, complex protocols that demand operator expertise, and time-consuming processes which often require manual monitoring over lengthy periods of up to 24 weeks.

While visual observation offers a simpler and more cost-effective alternative, it still depends heavily upon repeated checks by human operators. To address this challenge, innovative solutions such as time-lapse photography and video analysis have emerged, allowing researchers to capture images over time under controlled conditions and to crop them into targeted segments for analysis.

Nevertheless, two major hurdles remain: the training of analysts to accurately interpret these images and develop robust instability detection methods, and the complexity inherent in the detection of instabilities. In response to these challenges, Procter & Gamble (P&G) have approached laboratories with requests to enhance their equipment's performance in these two areas. In response, laboratories proposed the development of a deep-learning-based solution to execute necessary operations, leveraging convolutional neural networks renowned for their effectiveness in image analysis due to their ability to recognize spatial relationships within images. Convolutional neural networks are already transforming fields such as medicine, where they are seeing increased usage for both the detection and prediction of various diseases, particularly through early diagnoses based on images obtained from diagnostic equipment. These networks are applied for object detection within images, including identifying bubbles or droplets in multi-phase systems or detecting defects. The issue at hand can now be viewed as a defect-detection challenge, where instabilities manifest as defects in images of formulated liquid samples.

Here, primary research goals have included the development of a labelled dataset to establish image ground-truth using data from the P&G Research Centre, the selection of an appropriate deep learning architecture, and the ultimate development of an automated instability detector capable of timely identification, which would bear significant industrial relevance.

At the P&G Brussels Innovation Centre, over 6 000 time series of visual observations of fabric softeners were collected. Each time series was segmented into frames, from which a subset of 10 frames, referred to as a mosaic, was extracted and labelled by seasoned industrial experts. These experts categorized each frame as either stable or unstable, with the option to mark the specific types of instability present within the images and to select multiple instabilities where applicable. The dataset comprised 53000 labels across 45000 images, with certain mosaics being evaluated by multiple experts for labelling.

A major challenge here emerged from the imbalanced nature of the dataset: only 10% of images exhibited instability, while 73% demonstrated stable behaviour and 17% contained errors such as misaligned crops or empty frames due to the removal of vials absent notification. Such imbalances represent a common challenge in industrial machine learning applications, and to address this, research has evaluated multiple approaches, including training CNNs from scratch, utilising local binary patterns and scaling binary descriptors, random forest methodologies, applying transfer learning, and using geometric and colour augmentations to improve generalization, with confusion matrices here representing critical analytical tools. Ultimately, researchers established an ensemble model comprising eight individual models through a widely used strategy to enhance performance by combining the predictions of multiple networks. This model has, to date, yielded promising results.

To ensure the interpretability of the system, research here utilized gradient class activation mapping, which leverages the final layer prior to flattening in order to generate a heat map that highlights areas of stability within the image. This tool proved invaluable for explaining model decisions, particularly when addressing stakeholder scepticism.

Another notable research challenge in this area was the sheer volume of labelling material required, with thousands of images per time-lapse video requiring labelling. To mitigate this challenge, researchers utilized a mosaic approach, employing ten-frame time-lapse videos to train the detector and to subsequently analyse new samples within the context of the full time-lapse video history. Nevertheless, ongoing efforts in labelling and continuous model improvement remain essential, as the quality of outcomes is directly tied to the quality of underlying data.

The resulting automated detection system significantly improves the identification of instability in formulated liquids. As part of this effort, researchers have created a more refined and high-quality dataset, humorously named the "Fluid" dataset, for weight and liquid instability, which is now freely accessible on GitHub and offers a valuable resource for the broader research community.

DISCUSSION

The issue of solubility remains a challenging and multifaceted topic, often resisting straightforward characterization. While current research has explored solubility from a thermodynamic perspective, alternative methods also exist. Notably, several machine learning studies have approached solubility prediction using predefined molecular descriptors, such as high-energy running capabilities, polarizability, and polarity. Current research, however, aims to enable neural networks to autonomously identify relevant features, freeing interpretation from user assumptions, although drawbacks here include the necessity of significantly large datasets. To this end, researchers have integrated hybrid modelling and thermodynamics, among other approaches, to ensure sufficient data availability.

The prediction of solubility across various crystal structures and their diffusivity represents a promising direction for future research. While recent advances in quantum chemistry calculations and methodological improvements are progressively enabling the estimation of enthalpies for various polymorphs and crystalline forms, data availability remains the chief limiting factor. Expanding high-quality datasets will hence be critical for advancing predictive models in this field.

The application of transfer learning in solubility prediction currently lacks an established baseline framework universally adopted across the field, and the development of such a baseline is context-dependent. In thermodynamics and liquid phase property prediction, multitask models that simultaneously predict multiple related properties have demonstrated particular promise. Incorporating predictions of correlated properties can enhance the accuracy of specific property predictions while supporting the development of generalizable pre-trained models. Hence, as standardisation and collaboration increases, the establishment of shared baselines is anticipated to become increasingly feasible.

⦿ CAN AI ASSIST IN THE CREATIVE WORK PROCESS?

AI is increasingly being recognised as a positive transformative force across industries, including in energy and chemical engineering. While academic research has achieved remarkable advancements in AI capabilities, the industrial sector is still in the early stages of exploring its full potential. Within engineering firms, particularly those working in oil, gas, and energy transmission, large-scale projects progress from process design to construction stages which often involve billion-dollar investments. Despite the repetitive and structured nature of workflows involved, which include process simulations, PFD development, and the creation of P&IDs, the level of automation in these tasks currently remains of limited nature. One of the most creative and challenging aspects of industrial workflows is the generation of PFDs. The future research vision is to develop systems that, upon receiving client input, can automatically generate flow schemes, link these with simulations, and integrate them with data sheets to support PFD development. Such systems would align simulation outputs with design documents, streamlining the early stages of project development. While these capabilities are currently technically feasible, the practical solutions currently available in the industry lag behind the advanced approaches explored in academic research.

With respect to AI's potential to support process synthesis and design within the chemical and process industries, among the tools currently available is a process synthesis software package that consolidates over two centuries of industrial knowledge, primarily sourced from the German process sectors. Access to this is available upon request, and one such example is provided by the Process Design Centre in Breda, Netherlands, which offers specialized support upon request.

Despite the promise of AI tools such as language models in generating process flow diagrams for chemical synthesis, practical limitations remain burdensome. AI output is often inaccurate, and this observation reflects a broader trend observed in industrial settings, where researchers begin by leveraging best practices, literature, and existing resources to conduct initial experiments, while the transition from laboratory to industrial scale demands more nuanced and innovative approaches.

Engineers play a critical role in evaluating laboratory results, especially when the potential exists to recover raw materials effectively. At this stage, focus shifts to designing viable production processes. Unlike simple scale-up, industrial translation requires a rethinking of unit operations, often diverging significantly in nature from lab procedures. Here, creativity and experience become vital, with decisions to be made regarding which unit operations to employ and how these shall be executed. AI may here contribute valuable insights, although a degree of healthy scepticism remains warranted while the field remains in its early evolutionary stages.

One of the key challenges in scaling up involves the identification and monitoring of KPIs across distinct production scales. This process also necessitates accurate sizing and the understanding that scaling often entails technological transformation, as techniques that function efficiently at laboratory scale may not translate directly to larger systems. AI may here serve as a valuable support tool by aiding in the identification of discrepancies as well as in the proposal of adaptations.

Nevertheless, the application of AI should be framed as an initial step in the broader creative process. While AI tools may suggest technically feasible plant configurations, real-world implementation requires careful consideration of operational realities, including cleaning intervals, lifecycle considerations, and maintenance requirements. It is for experienced human engineers to understand such factors and to integrate these into the equipment selection criteria, while penalizing designs that fall short and fail to meet practical standards.

○ SHOULD AI BE INTEGRATED INTO CHEMICAL ENGINEERING CURRICULA?

As the role of AI expands within industrial sectors, a fundamental question arises: should AI and data science be integrated into the core curriculum of chemical engineering programs? Consensus among academic and industrial stakeholders points to an urgent need to modernize education to better align with contemporary digital demands. Experts acknowledge that their academic training previously lacked data science components, and bridging the gap between industry and academia is increasingly viewed as a critical endeavour in the field. With AI observing growing relevancy in the sector, as well as the skills required to leverage artificial intelligence in addressing upcoming challenges, future engineers must be prepared to engage with these technologies effectively.

To achieve this aim, educators emphasize the need to provide students with sufficient foundational knowledge to meaningfully utilize AI tools. For instance, the chemical engineering scale-up process involves a transition from laboratory experiments to full-scale industrial production, which often spans five to 10 years, with almost 50% of such transitions ending in failure. AI holds the potential to significantly reduce time-to-market and increase the probability of success, with 90% success metrics representing an ideal minimum, although engineers will have to be sufficiently trained to fully exploit such potentialities.

One proposed strategy here is the inclusion of collaborative, company-involved projects to engage students with real-world applications of machine learning and Python applications in chemical engineering contexts. Emphasizing practical experiences here enhances not only technical competencies but also fosters industry-relevant problem-solving skills.

A common thread in industry feedback is the importance of understanding how industrial machinery functions and how issues scale, as scaling up processes could represent one means of enhancing success rates while reducing lead times. For example, the disposal of insoluble residues may seem trivial at laboratory scale but can become critical in full-scale production, hence the need for laboratory approaches that simulate the complexities of real-world, large-scale operations.

Experts also recognize that current graduates often lack adequate education in data engineering. In sectors such as the pharmaceutical industry, digitization is advancing rapidly, with dedicated roles emerging focused on data quality and terminology standardization. Misinterpretations of terms, such as the dual meanings of “API” in computer science and pharmaceutical contexts, can undermine data integrity when interfacing with, for instance, SCADA systems. Educational programs must therefore teach not only programming and computer science, but also the standards and vocabularies necessary for effective data collection and management.

The shift from traditional languages such as MATLAB toward Python reflects broader industry trends, with Python now representing the language of standard within data analytics. Its inclusion in chemical engineering education is hence of paramount importance. However, data science should not be confined to a stand-alone subject. Instead, it should be integrated into the chemical engineering curriculum as a generalizable toolset, comparable to the case in the introduction of computers in engineering decades ago.

Of equal importance is the need for real data. While simulated data may serve as an entry point for training purposes, authentic industrial data is indefensible for producing valuable industry outcomes. Challenges persist in accessing high-quality data, however, and it is here of imperative nature that gaps be bridged with respect to data democratization.

Such considerations are particularly important with respect to SQL relational databases and MES platforms, which are often more complex than Pythonesque scripting environments, and commonly present a difficult challenge to engineers. This issue is becoming increasingly critical with the rise of cloud-based systems, rendering those in the field with knowledge of SQL at great advantage over their traditionally educated counterparts.

Regardless, not all engineers will need to become highly proficient in writing computer code. Instead, focus should lie on cultivating a general understanding of how AI tools work and how to apply various methodologies appropriately. Tools such as Copilot, GitHub, and ChatGPT may be able to execute code, but it is crucial that users know how to structure inquiries and interpret outputs effectively.

Elective courses in machine learning are already becoming more common in chemical engineering Master's programmes. However, uptake among students remains less than is necessary. To address this, educators suggest positioning machine learning not as a niche skill but as across-disciplinary enabler. To this end, resources such as the "Machine Learning for Chemical Engineers" GitHub repository offer promising entry points for students.

In addition to teaching AI use, academic institutions are urged to collaborate in developing open educational resources that keep pace with the rapid evolution of the field, in order to assist students in understanding the rationale behind AI's selection of specific models. From a statistical standpoint, particularly where pre-treatment data is concerned, experts strongly suggest incorporating AI tools, with leading process simulators such as Aspen now offering AI-enhanced features, including hybrid modelling techniques. Experts suggest that these should be incorporated into academic simulation training to keep pace with contextual AI progression. Some universities have already begun such implementation, with Nancy offering numerous elective courses in machine learning and related fields, and establishing committees to evaluate where AI efforts require further acceleration.

A broader pedagogical question also arises in this respect: should chemical engineers also be trained as data analysts, or should they simply be taught the language necessary to collaborate with data professionals? Consideration should here be given to the potential risk that students may struggle to find employment if their education leans too heavily on data science absent solid chemical engineering foundations.

Successful integration depends on multidisciplinary collaboration. While building advanced tools often requires dedicated machine learning or data engineers, chemical engineers bring irreplaceable domain knowledge to the table, particularly regarding the physical and chemical constraints of real-world processes.

In practise, researchers are increasingly engaged in the transition from proof-of-concept to the implementation of web application-based AI and mechanistic model-driven methodologies for process engineering, with companies increasingly expressing interest in generalist engineers who feature hybrid expertise in both chemical process engineering and data engineering, as well as in data analytics. As projects move toward production, the scale and complexity of data increases, requiring engineers to assume roles similar to those of product owners, in which they define objectives, identify user needs, and align technical outcomes with business goals. This trend underscores the crucial nature of comprehensive understanding on the part of graduates.

Ultimately, while education must prepare engineers to enter the workforce, it cannot adequately cover every possible future scenario. Lifelong learning is inevitable, and industry will certainly need to play a central role in supporting continuous development through on-the-job training programs. With the proliferation of tools and platforms available, educators are encouraged to harness these resources to offer more dynamic and adaptive learning environments.

● HOW CAN AI TECHNOLOGY WORK HAND-IN-HAND WITH HUMANS?

While AI continues to influence the landscape of publications concerning the advancements and future applications of machine learning tools in chemical engineering, existing literature is observed to lack a standardized approach. Frequently, data is accessible only upon request, or is in some instances entirely lacking, with corresponding code often being unavailable. A central question hence emerges: is it possible to establish a generalized standard procedure for drafting literature publications?

Although data on chemical reactions exists, it is not often structured in a way that supports AI applications, despite the time-intensive efforts spent generating it in laboratories. To avoid these datasets being locked behind corporate or proprietary interests, an open call has been announced to publish data in open-access databases. Furthermore, these open resources would enable standardized models to be benchmarked against validated datasets, rather than having engineers rely on questionable patent literature. Even when data is available upon request, journal policies may limit how this can be used.

For experimentalists, the public accessibility of data presents challenges. Frequently, when graduate students collect data, raw outputs from the analytical instruments are not immediately usable. Interpretation is what imbues data with its value. For example, although chromatograms can be shared, without contextual information on how they were obtained, they can easily be misinterpreted. Some researchers have adhered to open-access mandates by including all data directly in their papers, even if statements indicate availability upon request. Tools such as Origin can also allow data extraction from figures, but again, interpretation remains key.

In some academic settings, efforts are being made to define standardized vocabularies and guidelines for publication, particularly within newly launched PhD programs. These guidelines specify which data types should be included, either in primary texts or as supplementary material. Such an approach could facilitate the use of this data across laboratories and enhance the reproducibility of experiments. The question necessarily arises, therefore, as to whether a similar methodology could be applied to artificial intelligence and machine learning experiments.

There already exists a set of standards for various measurement types, along with recommended systems, such as those involving the testing of liquid extraction equipment. Experts suggest that it may be advantageous to explicitly publish these standards and to make them more visible and accessible. This need not be confined to scientific papers, and could perhaps be achieved through dedicated organizations, such as the EFCE, which could identify and promote specific methodologies.

As AI becomes more prevalent in laboratory settings, researchers observe the need to increase its user-friendliness. Many lab personnel are experts in their respective fields but avoid Python or command-line tools in favour of more familiar workflows. By embedding Python scripts into custom software, laboratories can develop internal tools with simple user interfaces that still leverage powerful AI capabilities. For instance, tools now exist with web-based interfaces that allow users to input molecular structures and receive solubility data, aimed specifically at experimentalists rather than AI specialists.

Just as users receive training before handling complex lab instruments such as NMR machines, experts recommend that similar training programs should be mandatory for those using AI tools, and including such tools and training as a requirement for publishing research could help ensure broader usage beyond the AI community.

Prior to the emergence of GenAI, emphasis focused primarily on explainability. Historically, many individuals expressed concerns regarding the functionality of these systems, often referring to them as "black boxes." Today, while such terms as "interpretability" and "explainability" are mentioned less frequently, the issue of trust in AI models remains just as relevant. To build trust, explainability should be embedded across the entire process, from data collection to machine learning model implementation.

There is also a need to incorporate the social dimension into study curricula. Chemical engineers interact with operators and technicians on the floor, and the success of the digital transformation depends on human factors. If end-users do not trust or understand the tools, even technically robust solutions may fail. Empowering users by giving them tools they can control and understand increases efficient engagement and reduces the risk of failures due to a lack of belief in the system. The social implications of such projects, particularly in change management, are significant: relevant terminology here does not represent an assemblage of mere buzzwords. That said, GenAI has already improved accessibility, allowing users to interact with systems via natural language or visual results, and these abilities have greatly enhanced user acceptance.

Nonetheless, data confidentiality remains a major barrier. For instance, in reactors using sensitive catalysts, data sharing may not be possible without exposing intellectual property. This raises the question: how can greater volumes of data be procured, and how can AI tools be applied in settings featuring only limited data? Here, techniques such as IPN modelling and other AI methods capable of functioning with sparse datasets represent potential solutions and are gaining traction, as researchers are increasingly assuming that significant data quantities may not always be available in industrial scenarios.

Explainability is especially critical in these contexts. Heat maps generated from the final layers of neural networks can help identify areas of importance and illustrate where significant effects occur, which enhances understanding and credibility. However, AI models and algorithms often identify patterns across multiple variables that may not be discernible in traditional 2D representations. These insights are still valuable, but may not always be easily explainable.

In process data analysis, the importance of explainability cannot be overstated. Many engineers in this field still prefer traditional tools such as Excel. For instance, an online monitoring platform named “Connection”, that includes low-code Principal Component Analysis (PCA) features, allows technical service engineers to input multiple variables and examine correlations. However, while these engineers can effectively use the tool, many struggle to explain its inner workings, such as the relevance of results on cumulative explained variance. Here, the main challenge to be addressed concerns how to effectively communicate data science concepts within the framework of low-code toolsets that engineers can readily use.

It would therefore be beneficial to establish a minimum knowledge baseline across the industry. A framework similar to Lean Six Sigma could help standardize the role of data science in the sector, with such standardization currently lacking. Regarding data sharing, despite technical potential, many organizations exhibit excessive caution regarding the dissemination of their data. Often, around 70% of data retained fails to generate revenue, and is hence being held without purpose. The obstacle here is not technical but political: any attempt to share data requires managerial approval, with managers typically declining such requests. This is also true of incident reporting, where similar permission must be sought. Regarding explainability, it is of crucial importance to address the issue of causality. Machine learning models have been developed that perform competently, yet they fail to adhere to causal principles, resulting in trends that lack physical validity and are thus rendered ineffective in practice. AI, therefore, can not be relied upon while it focuses on correlations rather than causations, yet this shortfall could be addressed by incorporating causality through causal models such as differential equations.

In many instances, a clear business case can be successful in convincing stakeholders and leadership of the benefits of data sharing and AI adoption. While reluctance may justifiably exist today, the long-term rewards of cooperation, to be seen five to ten years down the line, will become apparent as the community advances and AI models improve.

Academic-industry partnerships also play a crucial role here. Projects are increasingly being designed in which companies collaborate with research institutions on work that extends beyond private corporate interest, not necessarily disclosing their core issues directly, but offering proxy issues that can still lead to real solutions. Such approaches allow for innovation without requiring any exposure of sensitive data. In a similar vein, experts believe that there exist numerous opportunities for such innovative collaboration initiatives between industry and academia, through which companies could provide alternative datasets or specify the type of data they wish to collect, thereby facilitating meaningful research and development, again without exposing any sensitive data.

This collaborative model is already prevalent in AI-driven chemistry. In the broader field of AI research, establishing benchmarks is essential for the conducting of fair comparisons, and if companies were to collaborate through the sharing of even select flow-sheets or datasets, whether public, patented, or otherwise, training sets could be established that would serve as reliable reference points, allowing multiple parties to develop derivative models. When companies withhold data exclusively for internal use, issues arise concerning limitations on potential academic research in relevant areas. If academia is excluded from such discussions, future innovation in the field may be hindered. Conversely, such a challenge could also foster innovative ideas and creative solutions from diverse perspectives and disciplines.

ACKNOWLEDGMENTS

The publication of this white paper is an initiative of the SFGP and EFCE.
It was made possible thanks to:

François NICOL - SFGP President, France
Giorgio VERONESI - EFCE President, Italie

- the contributions from:

Mathieu CURA – Optimistik, France
Robert DAVID – Technord, Belgium
Maurizio DE MICCO - Univ.Naples Federico II, Italy
Anna JAEGGI - ETH Zurich, Switzerland
Ludovic MONTASTRUC - Toulouse-INP, France
Thibault NEVEUX – EDF, France
Artur SCHWEIDTMANN - Delft Univ. Tech., Netherlands
Mattia VALLERIO - Politecnico di Milano, Italy
Florence VERMEIRE, KU Leuven, Belgium
Massimiliano VILLONE - Univ. Naples Federico II, Italy

- the participants in the round-table:

Marco BASSETTO – LyondellBasell, Italy
Benoit CELSE – IFPEN, France
Gaia CHIARI - FH Münster - Dep Chem. Eng., Germany
Anton DE VYLDER - Covestro NV, Belgium
Maarten DOBBELAERE - Ghent University, Belgium
Vincent DUMOUILLA – Roquette, France
Arnesh PALANISAMY – Novartis, Switzerland
Daniel STARK – Axens, France
Nimet STERNEBERG - Clariant Produkte, Germany
Antonio VELARDO - Technip Energies, Italy
Verena WOLF-ZOELLNER - Montanuniversitaet Leoben, Austria

- the organizing committee:

Luigi Piero DI BONITO - Univ. Campania, Italy
Daniele MARCHISIO - Politecnico di Torino, Italy
Ludovic MONTASTRUC - Toulouse-INP, France
Martine POUX - EFCE – SFGP, France
Boelo SCHUUR - EFCE Scientific Vice-President, Netherlands
Kevin Van GEEM - Ghent University, Belgium

- the 130 attendees

APPENDIX

5th EUROPEAN FORUM ON NEW TECHNOLOGIES

ARTIFICIAL INTELLIGENCE in CHEMICAL ENGINEERING

Paris
13 December 2024

8:45 am • Welcome and Introduction
Giorgio Veronesi, EFCE President
François Nicol, SFGP President
Boelo Schuur, EFCE Scientific Vice-President

9:00 • Overview of artificial intelligence for process engineering

- ▶ Artificial intelligence & machine learning in the process industry: where we are and where we are going
- ▶ Different ways in which artificial intelligence can be used within the organization to optimize industrial processes
- ▶ Artificial intelligence tools for process modeling and simulation: a critical overview on potentialities and limitations so far

Questions and break

11:15 • Generative artificial intelligence and transformation in process engineering

- ▶ Transforming process engineering with generative artificial intelligence
- ▶ Process flowsheet generation by artificial intelligence: motivation & current state

Questions and lunch

13:30 • Artificial intelligence applications in chemical engineering

- ▶ Hybrid modelling and digital twin in clinker production
- ▶ Using machine learning for online 3D characterization of crystals in suspension
- ▶ Predicting solubility limits with machine learning
- ▶ Deep-learning methods for the image-based assessment of the physical stability of formulated liquids of industrial interest

Questions

15:45 • Round table

- 1 – Can AI help in a creative work process?
- 2 – Integrating AI into Chemical Engineering curricula
- 3 – How AI technology will be working hand-in-hand with humans?

17:00 • Closure



Société Française de Génie des Procédés
www.sfgp.asso.fr



EFCE
 European Federation of Chemical Engineering
 Fédération Européenne de Génie Chimique
www.efce.info

LOCATION

Hôtel MERCURE Gare Montparnasse
 40 rue du Commandant Mouchotte
 75014 PARIS – France
 Metro: 4, 6, 12, 13
 Stop: Montparnasse-Bienvenue

COORDINATOR

Martine Poux
 Martine.Poux@toulouse-inp.fr

CONTACT

Martine Poux
Société Française de Génie des Procédés
European Federation of Chemical Engineering

martine.poux@toulouse-inp.fr
<https://www.sfgp.asso.fr/>
<https://efce.info/>



September 2025 - SFGP production
ISBN - 978-2-910239-90-9

