Artificial intelligence in colloid and interface science: Current research, challenges and future directions

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in
Colloid & Interface Science

Artificial intelligence in colloid and interface science: Current research, challenges and future directions

Simha Sridharan[1,a], Tom Bailey[1,2,a], Agnese Marcato[3], Elena Simone[3] and Nicholas Watson[1,2]

Artificial intelligence (AI) and Machine learning (ML) are transforming colloid and interface science by enabling predictive modelling, autonomous experimentation, and accelerated material design. This review highlights recent advances organised in four topics: (1) prediction of basic physical properties; (2) image analysis; (3) process design, monitoring and optimisation; and (4) morphology and phase behaviour prediction. AI models have improved the prediction accuracy of interfacial tension, critical micelle concentration, foam stability, and complex structure–function relationships, in particular, integrated generative AI approaches support the design of new surfactants and emulsifiers. Image analysis has automated microstructural characterisation and enabled real-time quality control, while AI-enhanced process design has delivered digital twins, closed-loop optimisation, and sustainability-oriented workflows. Morphology and phase behaviour prediction has combined simulation-driven neural networks with generative approaches to accelerate material discovery. The future of AI applications in colloids will be shaped by experimental database design and standardisation, hybrid AI methods integrating physics and surrogate modelling, and AI agents leveraging large language models for literature mining, data curation, and experimental optimisation. Together, these developments promise to establish data-rich, physics informed, and increasingly autonomous research ecosystems for colloids and interface science, accelerating material understanding and design.

Addresses
[1] School of Food Science and Nutrition, University of Leeds, Leeds, UK
[2] National Alternative Protein Innovation Center (NAPIC), UK
[3] Department of Applied Science and Technology, Politecnico di Torino, Torino, Italy

Corresponding author: Watson, Nicholas (n.j.watson@leeds.ac.uk)
[a] First authors.

Current Opinion in Colloid & Interface Science 2025, **80**:101965

This review comes from a themed issue on **Food Colloids (2025)**

Edited by **Martin E. Leser**, **Brent Murray**, **Krassimir Velikov**, and **Stefan Salentinig**

For complete overview about the section, refer Food Colloids (2025)

https://doi.org/10.1016/j.cocis.2025.101965

Keywords
Artificial intelligence, Machine learning, Colloid science, Interface science.

Introduction

Colloids are multiphase materials (e.g. emulsions, foams, solid suspensions) ubiquitous in food, pharmaceutical, cosmetic and chemical sectors. The complexity of colloidal materials depends on several interconnected factors including: (1) their multiphase nature, (2) their structural complexity across different length scales, from the sub-nano to the macroscopic one, (3) the presence of multiple intermolecular interactions and microscopic effects (e.g., capillary forces) that are strongly related to pH, ionic strength, and temperature [1].

To understand the complex process-structure–function relationship of colloidal materials, experimental techniques to characterise their structure during processing, storage and use are combined with multiscale models [2]. Due to the multi-phase, multicomponent nature of colloidal systems, their characterisation often requires the use of multiple techniques (e.g., imaging, light scattering, zeta potential), generating a high volume of data. In this context the use of AI strategies can be extremely beneficial to organise, combine and extrapolate significant information from large datasets to better understand structural organisation and changes during the formation and performance of colloidal materials [3].

Artificial intelligence (AI) is a branch of computer science focused on creating systems capable of performing tasks that typically require human intelligence, such as pattern recognition, decision making, and problem solving. A key subset of AI is machine learning (ML), where supervised learning uses labelled data to train models for tasks such as classification and regression; whereas, unsupervised learning finds hidden structure in unlabelled datasets through clustering and dimensionality reduction. The recent rise of deep learning, based on multi-layer neural networks, has enabled breakthroughs in fields including image analysis, natural language processing, and scientific modelling [4]. 2023 was a breakout year for AI and ML with the launch of

Large Language Models (LLMs), a type of generative AI, trained on massive text corpora to capture linguistic patterns which output context-aware responses.

The general workflow for AI applications includes data collection, cleaning, and preprocessing, followed by feature selection. Models are then trained on subsets of data, validated to tune hyperparameters, and tested to evaluate performance before deployment [5]. Ongoing monitoring and retraining are often required as data or conditions can evolve.

AI offers key advantages, including the ability to capture complex nonlinear relationships, handle high-dimensional datasets, and automate time-consuming processes. These capabilities have enabled advances in sectors ranging from healthcare and finance, to materials science and food engineering However, limitations remain: AI models are often data-intensive, can act as "black boxes" with limited interpretability, may fail when applied outside their training domain, and often require substantial computational and environmental resources [6]. Addressing these challenges is an active area of research, with emphasis on explainable AI, transfer learning, and data-efficient modelling.

This review will focus on how AI has been used in colloids and interface science (section 2) through the following sub-sections: Basic physical property prediction (section 2.1); Image analysis (section 2.2); process design (section 2.3) and morphology prediction and phase behaviour (section 2.4). These sections will detail the key work in these areas and their current limitations. A final summary is also provided with the authors' thoughts on future directions of AI in colloids and interface science (section 3). A glossary of AI/ML terms which colloid and interface researchers may be unfamiliar with is available in supplementary material.

Current artificial intelligence activities in colloid and interface science

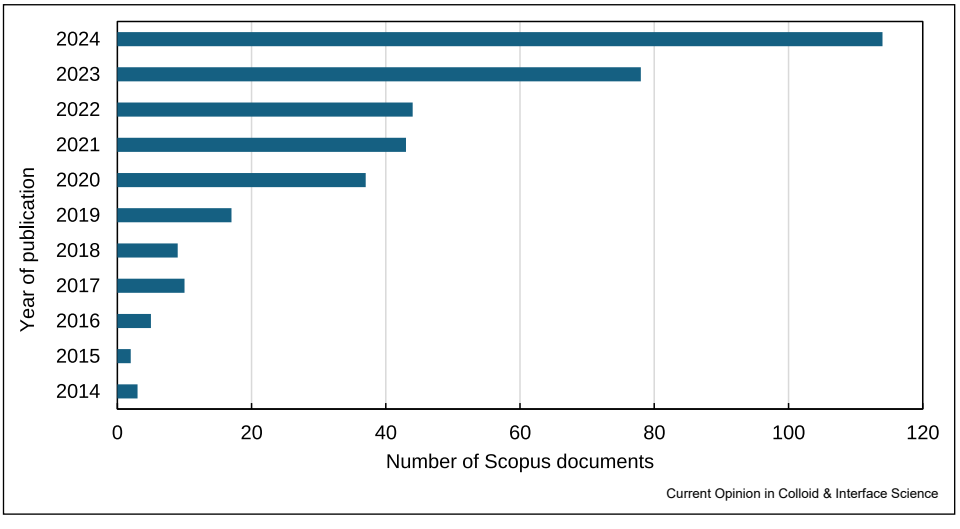
The first publications mentioning the use of AI and ML in colloids and interface science in their keywords or abstract started to appear in 2014 with increasing frequency from 2019 (Figure 1) across a broad range of topics (Figure 2).

Basic physical property prediction

The prediction of fundamental physical properties in colloids and interface science has seen rapid adoption of AI approaches, offering alternatives to first-principle models and empirical correlations.

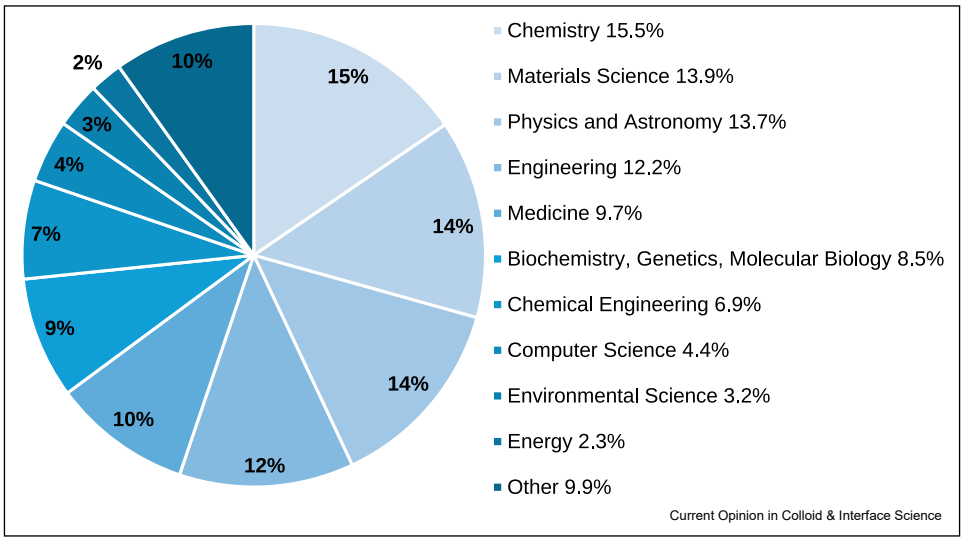
For interfacial and surface tension, several studies show promising predictive capabilities. Neural networks applied to experimental interfacial tension (IFT) data outperform conventional time-delay models based on prediction error metrics such as RMSE and rRMSE [7]. Similarly, boosted regression and other ML models predicting surface tension of hydrocarbon surfactants demonstrate strong accuracy but are limited to molecules structurally similar to the training set, highlighting the need for chemically diverse datasets and physics-based feature engineering [8]. Critical micelle concentration (CMC) prediction benefits from incorporating

Figure 1



Number of Scopus documents obtained by searching: (machine AND learning OR ai) AND colloid*. The search was performed on the 29th of July 2025. Of the 362 documents identified, around 74 are related to food (around 20 %).

Figure 2



Scopus documents obtained by searching: (machine AND learning OR AI) AND colloid* divided by topic. The search was performed on the 29th of July 2025.

structural descriptors and surfactant class information, and from modelling across multiple experimental conditions, representing a step beyond basic CMC models but still largely focused on sodium-based anionic surfactants [9,10].

More interpretable ML approaches are emerging, including 3D visualization of the relationship between surfactant structure and adsorption efficiency, which provides chemically meaningful insights but remains restricted to air–water interfaces [11], and feature attribution methods such as SHapley Additive exPlanations (SHAP) and Leave-One-Feature-Out (LOFO), which highlight structural drivers of surfactant clustering and adsorption.

Foam and rheological properties have also been investigated with the use of Machine Learning. For polysaccharide-based foams, generalized regression neural networks achieved high R^2 (>0.97) when predicting density, modulus, and foamability from minimal input variables, yet their complexity far exceeds what is warranted by the dataset, risking overfitting and making model generalisation and interpretability a challenge. This is a common issue in applied ML where simpler models are often ignored for more complex ones [12]. Rheological features have been modelled using classical rheological constants (e.g., Bingham, Casson models) as input features, showing how combining physical priors with ML can improve robustness [13]. Other applications, such as predicting viscosity in Konjac Glucomannan colloids, identified key microstructural features linked to storage modulus but were limited by small datasets lacking environmental parameters like pH and ionic concentration [14]. Several other studies suffer from similar issues of low sample size or narrow scope [15–17]. These examples illustrate a broader trend: high model complexity is often applied to limited data, with little consideration of overfitting or generalizability.

AI is also being used to design new molecules and colloidal architectures. Reinforcement learning (RL) frameworks combining graph neural networks and variational autoencoders allow bi-directional mapping between molecular structure and desired properties, creating a platform for targeted surfactant design [18]. RL has also been used to design novel protein nanomaterial assemblies, including icosahedral structures, demonstrating its potential for creating previously unobserved colloidal architectures [19]. Active learning methods improve efficiency in capturing complex molecular interactions [20], while neural-network-potential molecular dynamics simulations reveal new mechanisms in nanocluster dynamics [21]. These works illustrate a shift from descriptive modelling to generative design, although most rely on simulated datasets that do not yet capture full experimental complexity [22].

Clustering and unsupervised learning methods also appear, for example identifying salt concentration as a dominant factor controlling aggregation in elastin–polyethylenimine complexes [23], while consensus ML models, integrating neural networks with decision trees, predict microfluidic emulsion droplet sizes for unseen fluid–geometry combinations [24].

Gaussian process models within Bayesian frameworks demonstrate utility for sol–gel transition prediction in data-constrained conditions [25]. Although these approaches show conceptual breadth, the recurring issues of dataset scarcity, model interpretability, and lack of rigorous benchmarking persist [26].

AI shows strong potential for property prediction in colloids, from interfacial tension to inverse design. Yet limited data, lack of uncertainty quantification, and absence of open benchmarks are remaining challenges.

Image analysis

Machine learning (ML) applied to image analysis is advancing colloid and interface science by enabling automated characterisation and quality assessment.

A key area of focus is the use of image analysis to predict fundamental physical properties. For example, neural network-based models were combined with automated image feature extraction to determine the effective water diffusion coefficient in dairy powders, providing a high-throughput alternative to gravimetric measurements and improving understanding of microstructural effects on moisture transport [27]. Hybrid ML frameworks (combining ML with mechanistic understanding) incorporating image-derived particle descriptors have also been used to model particle formation kinetics, such as nucleation and growth, demonstrating computational advantages compared to purely mechanistic models [28].

Image classification for automated quality control is another major application. A convolutional neural network was used to classify emulsion images, delivering improved product quality evaluation compared to principal component discriminant analysis [29]. Complementary work applied random forest and multinomial logistic regression models to pharmaceutical emulsion images collected in-process, demonstrating that relatively simple, interpretable models can be effective for inline quality monitoring [30].

Image analysis is also being applied to study biological and chemical interactions. One study combined random forest, k-nearest neighbour, and support vector machine models with microscopy images to investigate biological effects of food additives and excipients, showing how image-derived features can be correlated with biological outcomes [31]. Other work focused on particle structure classification and state detection, including analysis of colloidal aggregation and breakup using deep learning and tracking morphological changes in emulsion systems [32,33].

Several studies have adopted more advanced or hybridised approaches. For example, one framework integrated image data with spectroscopy and textural metrics to enhance classification of heterogeneous systems [34]. Physics-informed models incorporating image features as constraints have been proposed to improve extrapolation beyond experimental training ranges [35]. Random forest methods have been applied for defect detection in manufacturing images, demonstrating that classical ML remains competitive, especially when labelled data are limited [36]. Deep learning methods have also been used to study surfactant adsorption and clustering behaviour [37], and advanced architectures such as vision transformers have been evaluated for colloidal assembly analysis [38]. These studies represent a shift towards multi-modal and transferable models, although all rely on curated datasets and significant expert involvement.

Integration of image analysis with surface chemistry models allowed accurate prediction of adsorption and wetting behaviour for novel materials [39], while ML-driven feature extraction from colloidal imaging datasets supported droplet coalescence and stability analysis in complex emulsions [40]. In polymer systems, ML has been used to map relationships between microstructural features and mechanical performance [41].

Despite successes, most image analysis models rely on small datasets with complex architectures, lack interpretability, and use bespoke workflows that hinder mechanistic insight, reproducibility, and transferability. Nevertheless, the reviewed literature highlights the transformative potential of ML image analysis for colloid and interface science. Applications now span property prediction [27], process modelling [28], product quality evaluation [29,30], biological interactions [31], adsorption phenomena [37] and material design [39–41].

Process design, monitoring and optimisation

Traditional process design often relies on empirical correlations and trial-and-error experimentation, which can be resource-intensive and slow to adapt to new materials and process conditions. ML provides a powerful alternative by learning directly from data, enabling the prediction of process performance, optimisation of operating conditions, and real-time adaptation based on sensor inputs.

Several studies demonstrate the use of ML to predict and optimise nanoparticle production. Ensemble models, including random forest, decision trees, and extra trees have been used to design and analyse continuous chitosan nanoparticle production processes, predicting particle size and polydispersity index with high accuracy compared to traditional design-of-experiment methods [42]. Feed-forward neural networks have been integrated with hybrid process models to predict the kinetics of particle nucleation and

agglomeration, demonstrating the benefits of combining physical descriptors with data-driven models to improve generalisation [28]. Similarly, random forest and neural network models were applied to microfiltration of oil-in-water emulsions to predict critical flux, allowing for rapid assessment of membrane performance under varied feed and operating conditions [43].

ML is also being used for process monitoring. For example, histogram-based droplet detection combined with principal component analysis and machine learning classification enabled automated categorisation of emulsion quality into four distinct classes, providing a non-invasive method for inline image-based process monitoring [29]. Supervised ML was also used to measure solid fat content during oleogelation of cocoa butter-based oleogels, in this case pulsed acoustic spectroscopy, temperature, and cocoa butter concentration were used as input to a gaussian process regression model [44].

Similarly, ML-based techno-functional property prediction of ingredients, using spline regression, random forest, and neural networks, was applied to food formulation and processing, allowing direct estimation of gelling, foaming, and emulsifying properties without extensive experimental characterisation [45]. These studies illustrate the potential for integrating image and sensor data into process design, though both noted challenges associated with dataset quality, manual curation, and variability in image acquisition conditions.

An ML-guided approach to predict properties of polymer composites demonstrated that combining experimental and computational datasets could accelerate process optimisation [46]. A curated open dataset supporting colloid and interface research was also introduced, enabling improved ML model training and benchmarking [47]. These resources support the development of generalisable models but remain rare and most studies rely on proprietary datasets, limiting reproducibility.

Hybrid models and physics-informed learning approaches are a growing modelling method for complex processes. For example, an integrated dataset and hybrid ML model were used for interface property prediction relevant to material design [48], while deep learning was used to study material assembly processes where traditional analytical models are insufficient [49]. These approaches demonstrate improved generalisation and physical interpretability compared with purely black-box models.

At the molecular scale, ML models have been used to explore interactions relevant to process design. Examples include modelling interaction potentials from quantum chemistry and molecular dynamics simulations [50] and applying neural networks to predict nanoscale assembly pathways for advanced materials [51]. ML has also been used to optimise wastewater treatment processes, demonstrating broad applicability beyond traditional colloid systems and membrane separations [52,53].

Real-time process monitoring is another emerging focus. ML models have been used for droplet stability analysis and microstructure classification in emulsions and for adsorption behaviour at fluid interfaces [54,55]. Clustering and unsupervised learning have revealed key factors influencing aggregation in elastin–polyethylenimine complexes [23] while Bayesian approaches have been applied to sol–gel transitions [56]. Surface adsorption and colloidal stability studies using explainable ML illustrate how interpretability can support process design [57]. ML is also enabling rapid prediction of nanoparticle thermodynamic properties and rheological behaviour [58,59], while consensus models have been developed for microfluidic emulsion droplet sizing across new geometries [24]. Other work has used Gaussian process models for uncertainty-aware predictions of colloidal systems [25], hybrid models for materials adsorption [60], and neural networks for spectroscopic feature–property correlations [61].

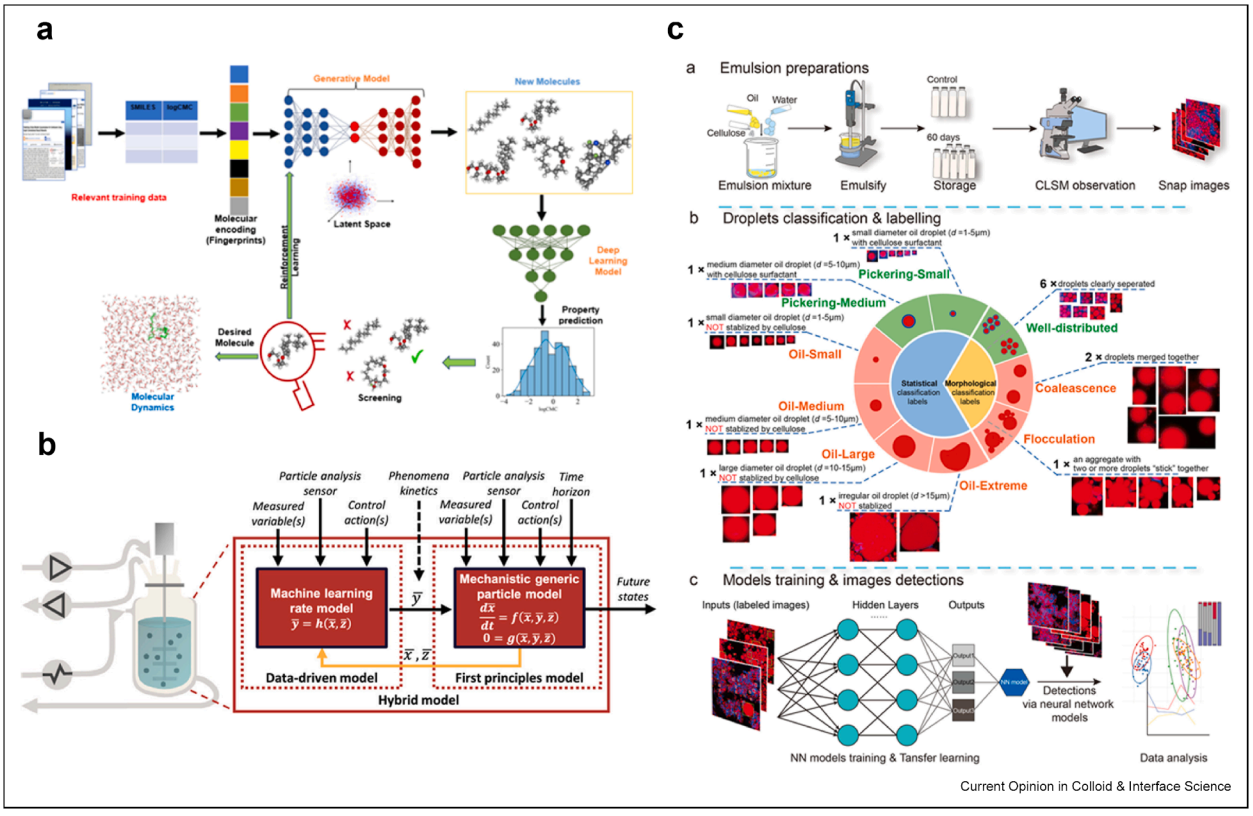
ML has demonstrated significant potential for improving process design, monitoring and optimisation in colloid and interface science, from nanoparticle synthesis [42] and filtration optimisation [43] to real-time monitoring [29] and advanced materials design [51]. Future work should focus on generating larger, open datasets, adopting interpretable and uncertainty-aware models, and embedding ML tools into digital process design frameworks.

Morphology prediction and phase behaviour

Predicting phase diagrams and morphological features is essential for understanding formulation stability, material performance, and processing efficiency, but experimental mapping and high-fidelity simulations can be costly and time-consuming. ML enables rapid exploration of compositional and processing spaces, identifying patterns linking molecular structure, thermodynamic conditions, and emergent morphologies.

Several works focus on directly predicting phase behaviour. For example, phase diagram mapping for multicomponent soft matter systems has been accelerated using ML methods capable of interpolating sparse experimental data and predicting untested compositions [26]. These approaches often employ classifiers or regressors such as random forests or support vector machines to classify regions of distinct phase behaviour, thereby reducing experimental workload. Similarly, an ML framework was developed to predict phase stability

Figure 3



Examples of different ML workflows used in colloids and interface science. (**a**) Overview of ML based workflow to incorporate reinforced deep learning to generate new surfactant molecules with ideal target properties design [18]; (**b**) Hybrid ML modeling approach to predict particle processes using mechanistic model and data-driven model [28]; (**c**) Image Analysis based Neural Network Workflow to predict properties of food emulsions [35]. All figures reproduced with permission.

and critical points for surfactant and polymer systems [62].

At the interface of morphology and phase behaviour, surrogate models have been used to capture nanoscale structural transitions. For example, active learning methods combined with Bayesian optimisation were integrated into self-driving laboratories to map morphological changes in block copolymers and nanoparticle assemblies [63]. These frameworks autonomously select experimental conditions predicted to yield novel structures, reducing human intervention and accelerating discovery. Similar ideas have been applied in molecular simulations, where neural network accelerate morphology prediction by learning interatomic forces from quantum calculations, enabling simulations at previously inaccessible scales [64].

ML has also been integrated with process simulation to predict morphology during complex operations. For example, ML-coupled computational fluid dynamics simulations were used to predict microstructure evolution in emulsification processes [65]. Experimental imaging data have also been linked to morphological descriptors using convolutional neural networks, enabling automated classification of colloidal phases and droplet size distributions [33].

ML methods have been applied to predict phase transformations under variable operating conditions. Hybrid physical–ML models have been developed to predict interfacial and phase behaviour during chemical processing and to optimise nanoparticle self-assembly [66,67]. A study on wastewater treatment highlighted the value of ML for complex multiphase phase separation processes [52], while ML-assisted crystallography has been used to predict phase structures in soft matter systems [68].

Emerging research has also emphasised explainability and hybrid modelling. A study combining ML with fundamental theories of adsorption and interfacial phenomena demonstrated improved performance and interpretability for predicting emulsion morphology

under varying formulations [35]. Similarly, ML-driven rheological property prediction supported improved understanding of morphological stability [13]. Innovations in experimental design, such as automated crystal structure [69], have been paired with ML to accelerate discovery of new phases and morphologies [70]. These examples illustrate that integrating ML with domain knowledge and automation enhances predictive power and physical insight.

In summary, ML has been applied successfully to predict phase behaviour [62], map morphological transitions [64], classify microstructures [33] and accelerate structural characterisation [69]. These applications demonstrate the transformative potential of ML, but also the need for broader datasets, improved interpretability, and integration into experimental and industrial workflows.

Summary, challenges and future directions of artificial intelligence in colloid and interface science

AI and particularly ML are rapidly transforming colloid and interface science, offering predictive power and automation for tasks historically dependent on empirical models and manual experimentation. However, challenges remain in dataset availability, model interpretability, and validation workflows. Given the breadth of AI and ML applications within the discipline standard modelling pipelines don't exists (different examples in Figure 3) making their use challenging for those lacking ML expertise.

The future of AI in colloid and interface science depends on overcoming critical data and integration challenges, many of which are already being addressed in other scientific domains.

First, experimental database design and standardisation are essential. Current datasets in colloids are often fragmented and lack harmonised metadata. Other fields demonstrate the power of standardisation: in materials science, the Materials Project and Open Catalyst Project have created structured, searchable datasets supporting AI-driven discovery of batteries and catalysts [71]. In biology, protein structure prediction by AlphaFold relied on highly curated and standardised protein databases [72]. Recent efforts have also produced standardized, multi-domain simulation datasets capturing complex dynamical systems, with unified metadata and output formats to support the development of first-principle models [73].

Similar curated, multi-technique datasets including both positive and negative results could accelerate colloid research; large language models (LLMs) already assist data curation and metadata harmonisation.

Second, hybrid AI modelling combining mechanistic (e.g. physics-based) approaches with data-driven methods will improve trust and industry adoption and ensure that AI integrates existing scientific knowledge and leverages the value of traditional modelling approaches. In molecular modelling, invariance to translation, rotation, and permutation is commonly enforced into graph neural networks [74]. In fluid dynamics, differentiable programming enables neural networks to learn flow behavior by minimizing the residuals of the Navier–Stokes equations during training [75] or can serve as a hard constraint of constitutive laws, boundary conditions and physical properties [76].

Third, AI-driven scientific workflows using LLMs and autonomous AI agents will streamline research pipelines. In drug discovery, AI agents such as IBM RXN and DeepMind's scientific assistant have automated literature review, experimental design, and synthesis planning [77]. In materials science, Bayesian optimisation combined with robotic labs has enabled autonomous closed-loop discovery of thin-film photovoltaics [78]. Similar AI-driven workflows could revolutionise colloid research, integrating literature review, automated data extraction, Bayesian experiment design, and robotic execution [79]. Pioneering efforts in earth system [80] and atomistic modelling [81] illustrate the potential of this approach to generalise across tasks, enable rapid adaptation, and community reuse.

These advances in other scientific fields illustrate how curated experimental databases, hybrid modelling, and autonomous AI agents can create a new research paradigm for colloid and interface science—one that is data-rich, mechanistically informed, and capable of accelerating discovery.

CRediT statement

Simha Sridharan: Investigation, formal analysis, writing – original draft. **Tom Bailey**: Investigation, formal analysis, writing – original draft. **Agnese Marcato**: Investigation, writing – review and editing. **Elena Simone:** Investigation, writing – original draft, writing - review and editing. **Nicholas Watson:** Conceptualization, Investigation, writing – original draft, writing - review and editing, project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

NW and TB acknowledge the UK National Alternative Protein Innovation Centre (NAPIC), which is an Innovation and Knowledge Centre funded by the

Biotechnology and Biological Sciences Research Council (BBSRC) and Innovate UK (Grant Ref: BB/Z516119/1). ES acknowledges the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 949229, CryForm) for funding. SS acknowledges the funding from UKRI Guarantee fund for Marie Curie Postdoctoral Fellowship (EP/Z000785/1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cocis.2025.101965.

Data availability

No data was used for the research described in the article.

References

1. Wong K, Arruda TR, Butler KT, Guldin S, Schrettl S: **Machine learning for food colloids: novel tools to advance fundamental understanding, stability, texture, and processability**. *Curr Opin Colloid Interface Sci* 2025, **78**, 101937, https://doi.org/10.1016/j.cocis.2025.101937.

2. Gołębiowski A, Buszewski B: **Characterization of colloidal particles of a biological and metallic nature**. *Microchem J* 2023, **191**, 108864, https://doi.org/10.1016/j.microc.2023.108864.

3. Parakhonskiy BV, Song J, Skirtach AG: **Machine learning in nanoarchitectonics**. In *Advances in colloid and interface science*. Elsevier B.V; 2025, https://doi.org/10.1016/j.cis.2025.103546.

4. LeCun Y, Bengio Y, Hinton G: **Deep learning**. *Nature* 2015, **521**: 436–444, https://doi.org/10.1038/nature14539.

5. Goodfellow I, Bengio Y, Courville A, Bengio Y: *Deep learning*, **1**. MIT press Cambridge; 2016.
This textbook is one of best references to understand the core principles of machine learning and deep learning.

6. Doshi-Velez F, Kim B: *Towards a rigorous science of interpretable machine learning*. 2017, https://doi.org/10.48550/arXiv.1702.08608. *ArXiv:1702.08608*.

7. Esteghlal S, Samadi SH, Hosseini SMH, Moosavi-Movahedi AA: **Identification of machine learning neural-network techniques for prediction of interfacial tension reduction by zein based colloidal particles**. *Mater Today Commun* 2023, **36**, 106546, https://doi.org/10.1016/j.mtcomm.2023.106546.

8. Seddon D, Müller EA, Cabral JT: **Machine learning hybrid approach for the prediction of surface tension profiles of hydrocarbon surfactants in aqueous solution**. *J Colloid Interface Sci* 2022, **625**:328–339, https://doi.org/10.1016/j.jcis.2022.06.034.

9. Abooali D, Soleimani R: **Structure-based modeling of critical micelle concentration (CMC) of anionic surfactants in brine using intelligent methods**. *Sci Rep* 2023, **13**, https://doi.org/10.1038/s41598-023-40466-1.
This study developed ML models to predict critical micelle concentration (CMC) of anionic surfactants under varying salinity, pH, and temperature. Stochastic gradient boosting achieved superior accuracy, highlighting AI's value for reliable CMC prediction across diverse conditions.

10. Chen J, Hou L, Nan J, Ni B, Dai W, Ge X: **Prediction of critical micelle concentration (CMC) of surfactants based on structural differentiation using machine learning**. *Colloids Surf A Physicochem Eng Asp* 2024, **703**, 135276, https://doi.org/10.1016/j.colsurfa.2024.135276.

11. Li S, Mao X, Cao X, Feng Y, Zhang Y, Yin H: **Visualizing molecular structure—Adsorption efficiency relationship through an interpretable machine learning strategy**. *Colloids Surf A Physicochem Eng Asp* 2025, **712**, 136400, https://doi.org/10.1016/j.colsurfa.2025.136400.

12. Ergün H, Ergün ME: **Modeling xanthan gum foam's material properties using machine learning methods**. *Polymers* 2024, **16**, https://doi.org/10.3390/polym16060740.

13. Al-Darweesh J, Aljawad MS, Tariq Z, Alajmei S, Yan B, Kamal MS: **Prediction of foam rheology models parameters utilizing machine learning tools**. *ACS Omega* 2024, **9**: 20397–20409, https://doi.org/10.1021/acsomega.4c00965.

14. Liu X, Hu L, Chen S, Ran Y, Pang J, Wu S: **Machine learning analysis for the rheological mechanism of polysaccharide colloids**. *J Mol Liq* 2025, **424**, 127093, https://doi.org/10.1016/j.molliq.2025.127093.

15. Muneer R, Hashmet MR, Pourafshary P, Shakeel M: **Unlocking the power of artificial intelligence: Accurate zeta potential prediction using machine learning**. *Nanomaterials* 2023, **13**, https://doi.org/10.3390/nano13071209.

16. Saddiqi HA, Ullah A, Javed Z, Ali QM, Jan MB, Ahmad I, Ahmad F: **Predictive modeling of emulsion stability and drop characteristics using machine learning: a study on surfactant influence and time dynamics**. *Food Bioprod Process* 2025, **149**:190–198, https://doi.org/10.1016/j.fbp.2024.11.019.

17. Sirovec S, Benković M, Valinger D, Sokač Cvetnić T, Gajdoš Kljusurić J, Jurinjak Tušek A, Jurina T: **Development of ANN models for prediction of physical and chemical characteristics of oil-in-aqueous plant extract emulsions using near-infrared spectroscopy**. *Chemosensors* 2023, **11**, https://doi.org/10.3390/chemosensors11050278.

18. Nnadili M, Okafor AN, Olayiwola T, Akinpelu D, Kumar R, Romagnoli JA: **Surfactant-specific AI-Driven molecular design: integrating generative models, predictive modeling, and reinforcement learning for tailored surfactant synthesis**. *Ind Eng Chem Res* 2024, **63**:6313–6324, https://doi.org/10.1021/acs.iecr.4c00401.
This study integrated generative AI, predictive modeling, and reinforcement learning to design surfactants, combining SELFIES, GNNs, and VAEs. It demonstrated accurate prediction and validation of molecular properties, underscoring AI's promise for tailored molecular design in colloid applications.

19. Lutz ID, Wang S, Norn C, Borst AJ, Zhao YT, Dosey A, Cao L, Li Z, Baek M, King NP, Ruohola-Baker H, Baker D: *Top-down design of protein nanomaterials with reinforcement learning*. 2022, https://doi.org/10.1101/2022.09.25.509419.

20. Deng Y, Fu S, Guo J, Xu X, Li H: **Anisotropic collective variables with machine learning potential for Ab initio crystallization of complex ceramics**. *ACS Nano* 2023, **17**:14099–14113, https://doi.org/10.1021/acsnano.3c04602.

21. Tiwari V, Karmakar T: **Atomistic details of nanocluster formation from machine-learned-potential-based simulations**. *Nano Lett* 2025, **25**:5940–5946, https://doi.org/10.1021/acs.nanolett.5c01340.

22. Long C, Liu S, Sun R, Lu J: **Impact of structural characteristics on thermal conductivity of foam structures revealed with machine learning**. *Comput Mater Sci* 2024, **237**, https://doi.org/10.1016/j.commatsci.2024.112898.
This study combined high-throughput simulations with ML to link foam microstructure to thermal conductivity, showing ANN models best captured complex dependencies. It highlights AI's role in designing thermally conductive foams for advanced thermal management applications.

23. Cobb JS, Engel A, Seale MA, Janorkar AV: **Machine learning to determine optimal conditions for controlling the size of elastin-based particles**. *Sci Rep* 2021, **11**, https://doi.org/10.1038/s41598-021-85601-y.

24. Lashkaripour A, McIntyre DP, Calhoun SGK, Krauth K, Densmore DM, Fordyce PM: **Design automation of microfluidic single and double emulsion droplets with machine learning**. *Nat Commun* 2024, **15**, https://doi.org/10.1038/s41467-023-44068-3.

25. Batra R, Joshi YM, Shanbhag S: **Navigating small datasets with machine learning: gaussian process modeling for**

colloidal gelation. *Langmuir* 2025, https://doi.org/10.1021/acs.langmuir.5c00754.

26. Jin SA, Kämäräinen T, Rinke P, Rojas OJ, Todorović M: **Machine learning as a tool to engineer microstructures: morphological prediction of tannin-based colloids using Bayesian surrogate models**. *MRS Bull* 2022, **47**:29–37, https://doi.org/10.1557/s43577-021-00183-4.

27. Maidannyk VA, Simonov Y, McCarthy NA, Ho QT: **Water effective diffusion coefficient in dairy powder calculated by digital image processing and through machine learning algorithms of CLSM micrographs**. *Foods* 2024, **13**, https://doi.org/10.3390/foods13010094.

28. Nielsen RF, Nazemzadeh N, Sillesen LW, Andersson MP, Gernaey KV, Mansouri SS: **Hybrid machine learning assisted modelling framework for particle processes**. *Comput Chem Eng* 2020, **140**, https://doi.org/10.1016/j.compchemeng.2020.106916.

29. Unnikrishnan S, Donovan J, MacPherson R, Tormey D: **An integrated histogram-based vision and machine-learning classification model for industrial emulsion processing**. *IEEE Trans Ind Inf* 2020, **16**:5948–5955, https://doi.org/10.1109/TII.2019.2959021.

30. Unnikrishnan S, Donovan J, Macpherson R, Tormey D: **In-process analysis of pharmaceutical emulsions using computer vision and artificial intelligence**. *Chem Eng Res Des* 2021, **166**:281–294, https://doi.org/10.1016/j.cherd.2020.12.010.

31. Reker D, Shi Y, Kirtane AR, Hess K, Zhong GJ, Crane E, Lin CH, Langer R, Traverso G: **Machine learning uncovers Food- and excipient-drug interactions**. *Cell Rep* 2020, **30**:3710–3716.e4, https://doi.org/10.1016/j.celrep.2020.02.094.

32. Altman LE, Grier DG: **Machine learning enables precise holographic characterization of colloidal materials in real time**. *Soft Matter* 2023, **19**:3002–3014, https://doi.org/10.1039/D2SM01283A.

33. Doan D, Echeveste DJ, Kulikowski J, Gu XW: **Machine learning analysis of self-assembled colloidal cones**. *Soft Matter* 2022, **18**:1532–1539, https://doi.org/10.1039/D1SM01466H.

34. Yao L, Ou Z, Luo B, Xu C, Chen Q: **Machine learning to reveal nanoparticle dynamics from liquid-phase TEM videos**. *ACS Cent Sci* 2020, **6**:1421–1430, https://doi.org/10.1021/acscentsci.0c00430.

35. Huang Z, Ni Y, Yu Q, Li J, Fan L, Eskin NAM: **Deep learning in food science: an insight in evaluating pickering emulsion properties by droplets classification and quantification via object detection algorithm**. *Adv Colloid Interface Sci* 2022, **304**, 102663, https://doi.org/10.1016/j.cis.2022.102663.

36. Patil A, Sægrov B, Panjwani B: **Advanced deep learning for dynamic emulsion stability measurement**. *Comput Chem Eng* 2022, **157**, 107614, https://doi.org/10.1016/j.compchemeng.2021.107614.

37. Mirzaei B, Nezamabadi-pour H, Raoof A, Derakhshani R: **An artificial intelligence image-based approach for colloid detection in saturated porous media**. *Colloids Surf A Physicochem Eng Asp* 2025, **713**, 136503, https://doi.org/10.1016/j.colsurfa.2025.136503.
This study introduced a lightweight deep CNN, enhanced by image dilation preprocessing, to detect colloids in porous media with greater accuracy than traditional methods, demonstrating AI's promise for monitoring contaminant transport and groundwater quality.

38. Schöttle M, Tran T, Oberhofer H, Retsch M: **Machine learning enabled image analysis of time-temperature sensing colloidal arrays**. *Adv Sci* 2023, **10**, 2205512, https://doi.org/10.1002/advs.202205512.

39. Sim JH, Yoo J, Lee ML, Han SH, Han SK, Lee JY, Yi SW, Nam J, Kim DS, Yang YS: **Deep learning model for cosmetic gel classification based on a short-time fourier transform and spectrogram**. *ACS Appl Mater Interfaces* 2024, **16**:25825–25835, https://doi.org/10.1021/acsami.4c03675.

40. Ayush K, Seth A, Patra TK: **nanoNET: machine learning platform for predicting nanoparticles distribution in a polymer matrix**. *Soft Matter* 2023, **19**:5502–5512, https://doi.org/10.1039/D3SM00567D.

41. Qu EZ, Jimenez AM, Kumar SK, Zhang K: **Quantifying nanoparticle assembly states in a polymer matrix through deep learning**. *Macromolecules* 2021, **54**:3034–3040, https://doi.org/10.1021/acs.macromol.0c02483.

42. Wu H, He J, Cheng H, Yang L, Park HJ, Li J: **Development and analysis of machine-learning guided flash nanoprecipitation (FNP) for continuous chitosan nanoparticles production**. *Int J Biol Macromol* 2022, **222**:1229–1237, https://doi.org/10.1016/j.ijbiomac.2022.09.202.

43. Tanudjaja HJ, Chew JW: **Application of machine learning-based models to understand and predict critical flux of oil-in-water emulsion in crossflow microfiltration**. *Industrial & Engineering Chemistry Research* 2022, **61**(24):8470–8477.

44. Metilli L, Morris L, Lazidis A, Marty-Terrade S, Holmes M, Povey M, Simone E: **Real-time monitoring of fat crystallization using pulsed acoustic spectroscopy and supervised machine learning**. *J Food Eng* 2022, **335**, https://doi.org/10.1016/j.jfoodeng.2022.111192.

45. Lie-Piang A, Hageman J, Vreenegoor I, van der Kolk K, de Leeuw S, van der Padt A, Boom R: **Quantifying techno-functional properties of ingredients from multiple crops using machine learning**. *Curr Res Food Sci* 2023, **7**, https://doi.org/10.1016/j.crfs.2023.100601.
This study used ML to predict techno-functional properties (gelation, viscosity, emulsion stability, foaming) of pea–lupine blends, showing cross-crop models are feasible though less accurate, highlighting AI's potential in sustainable formulation with minimally refined ingredients

46. Liu AL, Venkatesh R, McBride M, Reichmanis E, Meredith JC, Grover MA: **Small data machine learning: classification and prediction of poly(ethylene terephthalate) stabilizers using molecular descriptors**. *ACS Appl Polym Mater* 2020, **2**:5592–5601, https://doi.org/10.1021/acsapm.0c00921.

47. Chitre A, Querimit RCM, Rihm SD, Karan D, Zhu B, Wang K, Wang L, Hippalgaonkar K, Lapkin AA: **Accelerating formulation design via machine learning: generating a high-throughput shampoo formulations dataset**. *Sci Data* 2024, **11**, https://doi.org/10.1038/s41597-024-03573-w.

48. Gao H, Jia H, Dong J, Yang X, Li H, Ouyang D: **Integrated in silico formulation design of self-emulsifying drug delivery systems**. *Acta Pharm Sin B* 2021, **11**:3585–3594, https://doi.org/10.1016/j.apsb.2021.04.017.

49. Dijkstra M, Luijten E: **From predictive modelling to machine learning and reverse engineering of colloidal self-assembly**. *Nat Mater* 2021, **20**:762–773, https://doi.org/10.1038/s41563-021-01014-2. Nature Research.

50. Fong AY, Pellouchoud L, Davidson M, Walroth RC, Church C, Tcareva E, Wu L, Peterson K, Meredig B, Tassone CJ: **Utilization of machine learning to accelerate colloidal synthesis and discovery**. *J Chem Phys* 2021, **154**, 224201, https://doi.org/10.1063/5.0047385.

51. Epps RW, Bowen MS, Volk AA, Abdel-Latif K, Han S, Reyes KG, Amassian A, Abolhasani M: **Artificial chemist: an autonomous quantum dot synthesis bot**. *Adv Mater* 2020, **32**, 2001626, https://doi.org/10.1002/adma.202001626.

52. Lee WH, Park CY, Diaz D, Rodriguez KL, Chung J, Church J, Willner MR, Lundin JG, Paynter DM: **Predicting bilgewater emulsion stability by oil separation using image processing and machine learning**. *Water Res* 2022, **223**, 118977, https://doi.org/10.1016/j.watres.2022.118977.

53. Li H, Wang Y, Wang Y: **Machine learning for predicting the dynamic extraction of multiple substances by emulsion liquid membranes**. *Separ Purif Technol* 2023, **313**, 123458, https://doi.org/10.1016/j.seppur.2023.123458.

54. Elsoudy S, Akl S, Abdel-Rehim AA, Munyebvu N, Howes PD: **Design of experiments coupled with Bayesian optimisation for nanolubricant formulation**. *Colloids Surf A Physicochem Eng Asp* 2024, **693**, 134026, https://doi.org/10.1016/j.colsurfa.2024.134026.

55. Hafsa N, Rushd S, Alzoubi H, Al-Faiad M: **Accurate prediction of pressure losses using machine learning for the pipeline transportation of emulsions**. *Heliyon* 2024, **10**, e23591, https://doi.org/10.1016/j.heliyon.2023.e23591.

56. Esteki B, Masoomi M, Moosazadeh M, Yoo C: **Data-driven prediction of Janus/core–shell morphology in polymer particles: a machine-learning approach**. *Langmuir* 2023, **39**: 4943–4958, https://doi.org/10.1021/acs.langmuir.2c03355.

57. Li S, Mao X, Cao X, Feng Y, Zhang Y, Yin H: **Visualizing molecular structure—Adsorption efficiency relationship through an interpretable machine learning strategy**. *Colloids Surf A Physicochem Eng Asp* 2025, **712**, 136400, https://doi.org/10.1016/j.colsurfa.2025.136400.

58. Dulaney AR, Brady JF: **Machine learning for phase behavior in active matter systems**. *Soft Matter* 2021, **17**:6808–6816, https://doi.org/10.1039/d1sm00266j.

59. Leverant CJ, Harvey JA, Alam TM, Greathouse JA: **Machine learning self-diffusion prediction for Lennard-Jones fluids in pores**. *J Phys Chem C* 2021, **125**:25898–25906, https://doi.org/10.1021/acs.jpcc.1c08297.

60. Lavrentev FV, Rumyantsev IS, Ivanov AS, Shilovskikh VV, Orlova O Yu, Nikolaev KG, Andreeva DV, Skorb EV: **Soft hydrogel actuator for fast machine-learning-assisted bacteria detection**. *ACS Appl Mater Interfaces* 2022, **14**:7321–7328, https://doi.org/10.1021/acsami.1c22470.

61. Zhou Y, Ouyang Y, Zhang Y, Li Q, Wang J: **Machine learning assisted simulations of electrochemical interfaces: recent progress and challenges**. *J Phys Chem Lett* 2023, **14**: 2308–2316, https://doi.org/10.1021/acs.jpclett.2c03288.

62. Thacker JCR, Bray DJ, Warren PB, Anderson RL: **Can machine learning predict the phase behavior of surfactants?** *J Phys Chem B* 2023, **127**:3711–3727, https://doi.org/10.1021/acs.jpcb.2c08232.
This study applied ML to predict surfactant phase behaviour from limited data, showing classifiers can impute missing points but struggle with de novo diagrams due to bias, emphasizing the need for improved data coverage in phase behaviour prediction.

63. Dijkstra M, Luijten E: **From predictive modelling to machine learning and reverse engineering of colloidal self-assembly**. *Nat Mater* 2021, **20**:762–773, https://doi.org/10.1038/s41563-021-01014-2. Nature Research.

64. Campos-Villalobos G, Boattini E, Filion L, Dijkstra M: **Machine learning many-body potentials for colloidal systems**. *J Chem Phys* 2021, **155**, 174902, https://doi.org/10.1063/5.0063377.

65. Aslannejad H, Samari-Kermani M, Mohammad Nezami H, Jafari S, Raoof A: **Application of machine learning in colloids transport in porous media studies: Lattice Boltzmann simulation results as training data**. *Chem Eng Sci* 2022, **253**, 117548, https://doi.org/10.1016/j.ces.2022.117548.

66. Kameya K, Ogata H, Sakoda K, Takeda M, Kameya Y: **Prediction of particle agglomeration during nanocolloid drying using machine learning and reduced-order modeling**. *Chem Eng Sci* 2024, **294**, 120097, https://doi.org/10.1016/j.ces.2024.120097.

67. Kombo DC, Stepp JD, Lim S, Elshorst B, Li Y, Cato L, Shomali M, Fink D, LaMarche MJ: **Predictions of colloidal molecular aggregation using AI/ML models**. *ACS Omega* 2024, **9**: 28691–28706, https://doi.org/10.1021/acsomega.4c02886.

68. Kooti G, Dabir B, Butscher C, Taherdangkoo R: **A constrained machine learning surrogate model to predict the distribution of water-in-oil emulsions in electrostatic fields**. *Sci Rep* 2024, **14**, https://doi.org/10.1038/s41598-024-61535-z.

69. Ding L, Chen Y, Do C: **Machine-learning-informed scattering correlation analysis of sheared colloids**. *J Appl Crystallogr* 2025, **58**:992–999, https://doi.org/10.1107/S1600576725003280.

70. Meinders MBJ, Yang J, van der Linden E: **Application of physics encoded neural networks to improve predictability of properties of complex multi-scale systems**. *Sci Rep* 2024, **14**, https://doi.org/10.1038/s41598-024-65304-w.
This study showed physics-informed neural networks (PeNNs) outperform standard NNs in predicting emulsion viscosity under shear, excelling at extrapolation with limited data, underscoring the value of hybrid (ML + physics) modelling for complex colloidal systems.

71. Chanussot L, Das A, Goyal S, Lavril T, Shuaibi M, Riviere M, Tran K, Heras-Domingo J, Ho C, Hu W: **Open catalyst 2020 (OC20) dataset and community challenges**. *ACS Catal* 2021, **11**:6059–6072, https://doi.org/10.1021/acscatal.0c04525.

72. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583–589, https://doi.org/10.1038/s41586-021-03819-2.

73. Ohana R, McCabe M, Meyer L, Morel R, Agocs FJ, Beneitez M, Berger M, Burkhart B, Burns K, Dalziel SB, Fielding DB, Fortunato D, Goldberg JA, Hirashima K, Jiang Y-F, Kerswell RR, Maddu S, Miller J, Mukhopadhyay P, Ho S: *The well: a large-scale collection of diverse physics simulations for machine learning*. 2025. http://arxiv.org/abs/2412.00568.

74. Lubbers N, Smith JS, Barros K: **Hierarchical modeling of molecular energies using a deep neural network**. *J Chem Phys* 2018, **148**, 241715, https://doi.org/10.1063/1.5011181.

75. Raissi M, Perdikaris P, Karniadakis GE: *Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*. 2018. https://www.sciencedirect.com/science/article/pii/S0021999118307125.

76. Mohan AT, Lubbers N, Chertkov M, Livescu D: **Embedding hard physical constraints in neural network coarse-graining of three-dimensional turbulence**. *Phys Rev Fluids* 2023, **8**, 14604, https://doi.org/10.1103/PhysRevFluids.8.014604.

77. Schwaller P, Probst D, Vaucher AC, Nair VH, Kreutter D, Laino T, Reymond J-L: **Mapping the space of chemical reactions using attention-based neural networks**. *Nat Mach Intell* 2021, **3**: 144–152, https://doi.org/10.1038/s42256-020-00284-w.

78. Macleod BP, Parlane FGL, Morrissey TD, Häse F, Roch LM, Dettelbach KE, Moreira R, Yunker LPE, Rooney MB, Deeth JR, Lai V, Ng GJ, Situ H, Zhang RH, Elliott MS, Haley TH, Dvorak DJ, Aspuru-Guzik A, Hein JE, Berlinguette CP: *Self-driving laboratory for accelerated discovery of thin-film materials*. 2020. https://www.science.org.

79. Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, Bernstein MS, Bohg J, Bosselut A, Brunskill E, Brynjolfsson E, Buch S, Card D, Castellon R, Chatterji N, Chen A, Creel K, Davis JQ, Demszky D, … Liang P: *On the opportunities and risks of foundation models*. 2022. http://arxiv.org/abs/2108.07258.

80. Bodnar C, Bruinsma WP, Lucic A, Stanley M, Allen A, Brandstetter J, Garvan P, Riechert M, Weyn JA, Dong H, Gupta JK, Thambiratnam K, Archibald AT, Wu C-C, Heider E, Welling M, Turner RE, Perdikaris P: **A foundation model for the Earth system**. *Nature* 2025, **641**:1180–1187, https://doi.org/10.1038/s41586-025-09005-y.

81. Batatia I, Benner P, Chiang Y, Elena AM, Kovács DP, Riebesell J, Advincula XR, Asta M, Avaylon M, Baldwin WJ, Berger F, Bernstein N, Bhowmik A, Blau SM, Cărare V, Darby JP, De S, Della Pia F, Deringer VL, Csányi G: *A foundation model for atomistic materials chemistry*. 2024. http://arxiv.org/abs/2401.00096.