www.stemmaresearch.com ISSN (Online): Applied
STEMMA International Research Journal
(International Open Access, Peer-reviewed & Refereed Journal)
(Multidisciplinary, Bimonthly, Multilanguage)
Volume: 1 Issue: 1 September-October 2025

Role of Artificial Intelligence in Theoretical Modeling of Molecular Interactions in Binary Liquid Systems

Saurabh Kumar

Research Scholar, Department of Physics, Bareilly College, Bareilly (U.P)
E-mail ID- skverma8088@gmail.com

Abstract-

Artificial Intelligence (AI) is increasingly pivotal in theoretical modeling of molecular interactions in binary liquid systems, addressing challenges in accurately predicting complex phenomena such as phase separation, nonideal mixing thermodynamics, and solute–solvent interactions. AI's role spans generating predictive potential energy surfaces, applying generative and inverse modeling for configuration space exploration, and enhancing sampling efficiency through active learning. Neural network potentials and graph-based molecular representations facilitate modeling of multicomponent interactions with improved accuracy and transferability. AI-driven multiscale coupling frameworks link quantum, molecular, and continuum scales, promoting comprehensive simulation of binary mixtures. Critical evaluation of AI models focuses on thermodynamic consistency, conservation laws, and transferability across varying compositions and thermodynamic conditions. Data curation, benchmarking, and transparency underpin reproducibility and trustworthiness, while computational efficiency and hardware utilization remain crucial for practical deployment. AI advances enable hypothesis formulation and design of experiments in binary liquid mixtures, paving pathways for accelerated discovery and enhanced understanding of molecular interactions. The integration of AI with classical thermodynamics and molecular theory offers promising avenues for advancing chemical sciences in predicting and controlling complex liquid behaviors.

Keywords: Artificial Intelligence, Binary Liquid Systems, Molecular Interactions, Neural Network Potentials, Generative Modeling, Thermodynamic Consistency, Multiscale Simulation.

Introduction

Binary mixtures constitute a fundamental, widely studied class of liquid systems, central to both theoretical development and applications. Phenomena such as phase separation and nonideal mixing thermodynamics generate considerable interest. Extensive molecular simulation studies and many-body theoretical approaches explore mixing, nonideal solvation, and solvent effects on solute reactions, yet gaps remain. Strong microscopic coupling, complex interfacial phenomena, and solute–solvent interactions challenge the accurate and efficient modeling of binary systems. AI

approaches hold promise in molecular modeling; various paradigms, domains, and proposals target specific challenges. The adoption of diffusion generative models for configuration-space exploration lacks physico-chemical content; presampling, accurate probability conditioning, and proxy energies mitigate costly many-body evaluations. Several multi-fidelity, AI-accelerated, and transfer-learning schemes offer improved prediction, data-efficiency calibration, and generalization across chemical-identity perturbations. AI-driven, two-dimensional (2D) representations, extended to three-dimensional (3D) mixtures, encapsulate symmetry and invariance, simplifying elaborated transfer-learning schemas. Graph approaches excel in the inclusion of soft constraints—complex topological, molecular graphs cater to extensive solute–mixture multi-composition, transfer-learning requirements.

AI methodologies for binary liquids span a diverse spectrum: facilitating datum synthesis, iterative sampling enhancement, and broad-ranging three-scale coupling interfacing. AI-guided aqueous-organic models exploit experimental proximity, enabling energetics guidance across various compounds; supercritical fluid modulation toward low-density regimes incidentally circumscribes high-temperature, low-density arrangements. LIW systems profiting from joint density-function-investigated energetics unveil restricted, explicit subgraph, yet under-sampled accelerants necessitate action by design. Tracers elucidate component interplay through AI-detected minute, yet crucial, interaction sparsity—current integrated frameworks retain solute-unaware inadequacies, underscoring further efforts within multiscale boundary.

Proxy models remain instrumental in refining AI models, especially within highly accurate many-body-ground frameworks. The generalization-acquisition tension among composition, temperature, and pressure substantially hampers LIW advancement—AI formations adhering to applicable, transferable, boundary-responding, or scenario-adaptive bases offer potential directional enhancement.

Theoretical Framework for Binary Liquid Mixtures

Mixtures of two immiscible or partially miscible liquids, commonly known as liquid–liquid binary mixtures, form a crucial class of colloids. In these mixtures, substances dissimilar to the solvents are often dissolved to improve the solubility of specific compounds of interest or to enhance certain physicochemical properties. Well-established experiments exist for measuring the density, refractive index, and ultrasonic velocity of liquid mixtures; however, direct determination of the associated solute–solvent interactions remains exceedingly challenging. Theoretical models, particularly prediction equations based on the properties of pure components, have been developed to address such measurement difficulties for a variety of binary liquid mixtures, including nonideal-water-organic solute systems (G. Arturo, 2005). Experimental characterization of liquid mixtures is both cumbersome and infeasible for all candidate combinations, emphasizing the need for predictive models. Established activity-coefficient models—UNIFAC, COSMO-RS, and the NRTL activity-coefficient equation—enable representation of liquid mixtures by pure-component information, yet these approaches do not provide insight into liquid–liquid solvation interactions. With the advent of machine learning and artificial intelligence, new opportunities arise to gain insight into the elusive arch of multicomponent solvation through predictive modeling of solute–solvent interactions within binary mixtures.

Artificial Intelligence Paradigms in Molecular Modeling

Binary liquid mixtures are forms of matter composed of two distinct liquids whose phases can strongly interact. Designers of binary liquid-based media typically seek to

enhance favourable properties and limit undesirable ones. Aqueous-organic mixtures are ubiquitous, yet aqueous-ionic liquids are emerging as greener alternatives for dissolving non-polar species. Ensuring reliable access to physical properties remains crucial for advancing predictive modeling. Aqueous-organic systems span vast concentration ranges, making accurate initialization challenging; AI facilitates a complementary approach by modelling systems containing very dilute, non-volatile additives. Further, large numbers of archived and fresh datasets are already available from both experimentation and simulation, streamlining data driven method development and rapidly extending the range of system compositions, temperatures, and pressures while refining predictions of structure–property relationships (Jung et al., 2021) ; (Cao & Tian, 2021).

Supervised Learning for Potential Energy Surfaces: The potential energy surface (PES) arises in a variety of fields, from atomic-scale processes to geological time scales. It describes the interaction of inter-particle energies, where pairs of atoms share similar scales of a few eV in bonding atoms. Intermolecular interactions in many-body systems involve much higher numbers of particles but stay within the range of bond formation and breakage. Molecules diffuse among energy wells established by maximum substitutions to this basin and therefore can reach the thermodynamic stable state. Within these physical perspectives, many simulation or modelling approaches of molecules fusing experimental data of thermodynamic, partitioning parameters, transfer data obtained by physicochemical springboard from clothes-cleaning agents aimed at chemical detection of components of binary water–the solvent of choice in many molecular structures–organic systems, (Zaverkin & Kästner, 2021) the possible quantification of solvation structure of potential-energy-mapping surfaces either volumic or in terms of partitioning at two different pressure on an ionic liquid additive to mixture and thus the identification of any detail structure of trace non-ionic component resulted by the identification of solute within 1–8 waters in conventional – or lime to midrange and no-water-accompanied counting strict area on the mixture-within-a-mixing environments gives their own informative details at only simple presence level.

Artificial potential-energy-surface (PES) approximation or directly potential mapping is attainable without auxiliary messy automatic-setup huge-spectrum-dataset generation phase. Accurate molecules with physics–chemistry compliance screened from wide array of topologies, simple larger-format reinforcement estimation on occasion or huge optimization non-termination-enabled initial setting on huge acoustic aspect ligands such as graph–continuum–molecular interfaces renders only potential-energy directly-mapping-safe modeling based only on molec number, rather than atom number in many advanced quantum-chemical PES method.

Generative and Inverse Modeling for Configuration Space: Exploring the full configuration space of molecular interactions is a formidable challenge that emerges in various contexts. Generative and inverse modeling aim to alleviate this difficulty by learning a probability distribution over configurations or by generating configurations that meet predefined criteria. These principles apply to binary molecular interactions, where the goal is to sample mixtures whose properties closely align with those of experimental systems.

Two key paradigms have advanced generative and inverse modeling: probability modeling, which characterizes target distributions, and the use of learned objectives paired with an adjustable generative model, which facilitates the generation of compliant configurations (Herzog et al., 2022) ; Schwalbe-Koda & Gómez-Bombarelli, 2019). Probability models enable sampling from the approximate distribution while

varying the conditions associated with the samples. For example, a common configuration involves using embeddings of the relevant thermodynamic state as conditioning information or latent variables. In cases where defining a full probability distribution proves intractable, conditional approaches may still be effective. Generative classifiers and other similar strategies allow the use of non-conventional objective functions that are straightforward to specify and facilitate configuration generation in order to achieve predefined states.

The empirical assessment of these models often centers around the fidelity of their outputs, scrutinizing either properties indicative of the target distribution or targeted observables. The active selection of training datasets becomes paramount, encompassing the definitions of criteria governing a good configuration concerning the molecule and properties of interest. When the objective emphasizes properties such as viscosity or diffusion coefficients instead of matching ensemble distributions, strategies that maximize the anticipated property or align closely with the expected equation of state become vital (Schwalbe-Koda & Gómez-Bombarelli, 2019).

Uncertainty Quantification and Model Validation: The modeling of molecular interactions is critical for advancing chemical engineering and materials science. Self-consistent thermodynamic models of binary liquid mixtures remain elusive despite considerable advances in molecular modeling and theory; a gap between theory and experiment persists. Artificial intelligence (AI) can potentially constrain thermodynamic models of molecular mixtures, accommodating diverse interactions across many chemical species and states of matter. By leveraging large data sets of molecular configurations, interactions, and thermodynamic properties, AI elucidates governing principles and raises the prototype theory to a higher level of sophistication.

AI-driven approaches to liquid systems focus primarily on substantial portions of the Configuration Space (CS), often neglecting the tracking of uncertainty about molecular specifications (e.g., certain pair potentials among mixture species). It is desirable to complement sophisticated large-data training of models describing mechanical structure and bulk thermodynamics with systematic attention to the uncertainty surrounding the molecular specification that drives these macroscopic properties.

AI-Driven Methods for Binary Liquid Systems

Artificial Intelligence (AI) plays an instrumental role in the theoretical modeling of molecular interactions in binary liquid systems. It provides key algorithms to: (i) predict multicomponent interactions in heterogeneous systems; (ii) represent mixtures of varying composition under different environments; and (iii) select data points for model building with a minimal number of queries. The modeling of binary liquids presents several challenges for materials science and engineering. Combining an explicit representation of molecular species with a fully connected neural network allows knowledge about the structure of mixing environments to be integrated. The availability of computational frameworks for quantum-chemical calculations of solute–solvent interactions enables machine-learning models to be trained on relevant properties from straightforward ab initio calculations. Finally, multiscale approaches link the microstructural information encoded in molecular configurations to continuum models, bridging the gap between quantum and macroscopic phenomena. Mulitiscale coupling frameworks enable AI models to even be incorporated in quantum–continuum schemes.

Neural Network Potentials for Multicomponent Interactions: Neural network potentials (NNPs) have emerged as powerful tools for modeling multicomponent

interactions in complex molecular environments. To facilitate their applications in binary liquid mixtures, different NNP architectures have been put forward, focusing on the development of suitable training protocols and local and global descriptors. The first approach, utilizing a single global descriptor for state identification, enables efficient training on small data sets, while subsequent transferability tests demonstrate the model's applicability to other aqueous systems and permanent organic solutes and even broader concentrations. The importance of local structural features in multicomponent liquids has prompted the consideration of graph-based molecular representations. These frameworks, encompassing atomic and edge features and message-passing schemes, allow NNPs to predict the energies and forces associated with target components given mixtures of widely varying identity and concentration. Binary mixtures involving a nonionic trace component highlight the potential of AI to enhance understanding of subtle solute interactions overlooked by traditional models.

Graph-Based Representations of Molecular Networks: Molecular graphs capture chemical information from the connectivity of atoms and bonding arrangements. Nodes correspond to atoms; edges represent bonded or nonbonded interactions. Interacting particles in materials can also be described via graphs. For binary liquids comprising two substance types, graph-based representations can encode fixed connectivity with fully flexible position and orientation in a single unified framework.

Molecular systems can be described by graphs consisting of nodes connected by edges. In chemistry, the nodes typically represent atoms and the edges denote covalent bonds. For compounds composed of more than one species, edge types may differentiate between different bond types or atom types. In physical sciences, graphs can also represent the connectivity of interacting particles in the same material. An atomic graph can represent a molecular machine as a set of coupled mechanical parts or a solid as particles in a material design. Graphs enable reduced dimensions and long-range interactions to capture structural characteristics and physics and self-assembly of materials over extended length and time scales.

Liquid–liquid binary mixtures of an aqueous layer and a nonaqueous organic phase are ubiquitous in nature and industry. Popular examples include extraction of high-value organic compounds from water–organic waste streams and separation of biofuels, pharmaceutical products, perfumes, and food additives from aqueous fermentation broths. Nonionic trace components dissolved in dominant binary solvents influence solvation strength and selectivity on the dominant solute, leading to more desirable product crystallization, separation, and purification. Organic solutes inducing hidden hydrogen-bonding interactions universally appear in aqueous and ionic liquid systems, requiring numeric evidence. The spontaneous emergence of aqueous and organic layers from binary mixtures of chemically unlike components exemplifies sufficient heat transfer via solute condensation in seeding condensation to increase condensation rate of droplets.

Active Learning for Efficient Sampling: Active learning facilitates efficient sampling in interatomic potential modeling by guiding the selection of the most informative configurations for training. By iteratively querying input configurations expected to reduce uncertainty most, active learning accelerates high-throughput searches for new materials and enables de novo exploration of potential-energy surfaces with reduced overhead. Hierarchical sampling methods extend the strategy to model-free generation of generalized conductive, semiconductive, or other interatomic interactions across diverse materials. Such techniques are instrumental in constructing general-purpose interatomic potentials that significantly decrease computational

expenditures while preserving high fidelity in subsequent simulations. Active learning and hierarchical strategies have also been successfully applied in machine learning force-fields for complex covalent compounds, inorganic crystals, ionic liquids, metallic glasses, and various nanostructures.

The active learning process consists of sequentially sampling new atomic configurations, estimating the model's uncertainty, and selectively incorporating data to improve the model. The procedure is initiated from an initial dataset of density functional theory (DFT) calculations performed on disordered configurations. Molecular dynamics simulations are then executed using the best available models to explore atomic configurations, with uncertainty measures governing the inclusion of novel configurations for additional DFT calculations. The models are retrained with the entire assembly of data after each round. This iterative loop continues until the model captures relevant features of aluminium crystals, such as face-centred cubic (FCC), hexagonal close-packed (HCP), and body-centred cubic (BCC) structures. Notably, the strategy does not bias the sampling toward specific types of configurations, permitting the model to encounter a wide variety of structures naturally. The active learning algorithm exhibits a high degree of parallelisability; it can engage numerous computing nodes to generate approximately 6000 DFT calculations within a single round. The method effectively predicts equilibrium and kinetic properties of aluminium's crystal and liquid phases, including defect energies and liquid–solid coexistence curves. It is especially geared toward extreme, nonequilibrium processes, as successfully demonstrated in shock simulations involving 3 million atoms, for which force prediction errors remain comparable to typical force magnitudes (S. Smith et al., 2020).

Multiscale Coupling: From Quantum to Continuum: Theoretical modeling approaches at the quantum, molecular, and continuum levels are increasingly combined in rigorous multiscale coupling frameworks to link phenomena separated by orders of magnitude in time and length while remaining consistent with fundamental physics. Such approaches enhance computational efficiency while broadening applicability beyond conventional unresolved regimes (Karra et al., 2023). Multiscale methods involving quantum and molecular-mechanical degrees of freedom have recently been extended to connect molecular-dynamics trajectories to reactive-transport solutions at the continuum or mesoscopic level (Le Piane et al., 2020). These methods focus on systems in which charge density or ultrafast processes evolve on timescales above the molecular-dynamics cutoff and transport dominates subsequent relaxation. Multiscale approaches are also being enacted at the molecular and continuum scales of classical fluid mechanics, where retention of microscopic nonequilibrium phenomena is often critical to predictions. The projection of quantum-mechanical structures and dynamics onto classical fields is similarly being integrated into quantum-classical multiscale methods. Specific multiscale coupling strategies that incorporate neural-network potentials and gradient-embedded architectures for machined-learned-coupled schemes offer opportunities for AI-guided investigations of binary mixtures.

Model Evaluation and Physical Consistency

The quality and reliability of AI-generated molecular models depend on the evaluation of their physical consistency, which is critical for trustworthy predictions of system properties. Basic requirements for physical consistency include thermodynamic consistency, respect for conservation laws, and validation against experimental data; the applicability of these criteria for systems modeled in AI-driven research is examined in detail for binary liquids. Thermodynamic consistency implies convergence of equation-of-state predictions under limiting conditions, and several established equations of state are available for different binary mixtures. Respect for

conservation laws mandates that continuous physical processes must not induce jump discontinuities in the modeled potential energy surface; the Maxwell relations and eight equilibrium relations among excess thermodynamic quantities provide additional relevant criteria. Transferability is a key concern in AI-driven modeling of binary fluids; simple metrics are proposed for the quantitative assessment of transferability across compositions, temperatures, and pressures.

Thermodynamic Consistency and Conservation Laws: Molecular interactions in binary liquid systems underpin fundamental phenomena in chemistry and engineering. While they remain poorly understood, significant insights can be gained from theoretical modeling of the relevant potentials. Nevertheless, no widely accepted candidates exist due to their intricate nature, and density functional theory incurs prohibitive computational costs. Theoretical descriptions of binary mixtures therefore lag behind those for pure fluids. Artificial intelligence (AI) and data-driven methods are beginning to fill this gap: they provide empirical approximations to the equilibrium configurations and potential energy surfaces (PES) of binary mixtures, based on data collected from related systems. Diverse architectures and training approaches have been proposed, addressing specific substances, properties, and experimental conditions, and progressively enhancing the state of the art. Mapping the current landscape of AI techniques in this field facilitates the search for an adequate model, informs ongoing development, and advances both theoretical and practical understanding of fluid mixtures.

Robust theoretical modeling of fluid systems hinges on compliance with basic thermodynamic principles. Conservation of mass, momentum, and energy is imposed naturally in most AI methods by the mathematical structures of the employed models. Together with thermodynamic consistency, these principles guarantee that the predictions made for substances of relevant thermo physical properties constitute a reliable basis for extrapolation beyond the available data. Thermodynamic consistency ensures the fulfillment of fundamental equations relating bulk observables at equilibrium, such as the Gibbs–Duhem relation, the Clapeyron–Clausius equation, and the statements of the second law, on the PES of the mixture. It further requires the construction of corrected models when interspecies interactions need to be incorporated into pre-existing, pure-component representations from the literature. Once validated, these AI modeling approaches represent a powerful tool for theoretical modeling of binary mixtures.

Transferability Across Compositions and Conditions: The ability of models to extrapolate meaningful properties across chemical compositions or thermodynamic conditions is among their most powerful features. The models covered in this section were trained with mixtures at one set of mole fractions, temperatures, and pressures, yet predictions can be made for distinct compositions and conditions. This generalization capability holds tremendous practical significance. It enables the identification of mixtures where specific target properties meet predefined criteria, even when exploratory data about those mixtures are scarce. Having no prior knowledge of the new compositions, temperatures, or pressures, the AI-infused instructions continue to yield relevant physical insights.

Different measures characterize the transferability of models trained on chemical- or phase-space subsets. One fundamental metric quantifies the shift of equilibrium configurations under different compositions, taken as the average pairwise deviation between non-hydrogen atoms from the specific starting equilibrium structure. Generalization across branched and linear alcohol mixtures was examined using a second metric: the prediction of water–metal free-surface energies with models trained

on water–desolvated-organic components. Transferability across distinct conditions—specifically, wide variations in temperature and pressure—was also assessed, demonstrating that both substituent type (saturated vs. unsaturated) and substituent functionality (fewer vs. more H-bond donors) influence surface-energy predictions. As in the previous cases, the AI-guided methodology offered direct pointers to these systematic dependences, together with insight into the generalization characteristics of the models.

Computational Considerations and Best Practices

Rigor in data curation and management is essential for machine learning (ML) research across any domain. Benchmark datasets that are compact, representative, decoupled from biased datasets, and publicly available reduce variability in training setups and facilitate fair comparisons among different methods or models considered on a similar task. Rigor in generating benchmark datasets, as well as public accessibility of training codes and detailed specifications of training setups, enhances the reproducibility and transparency of ML research over the deep learning literature in the past 20 years. Improved reproducibility and transparency also allow research to benefit from accumulated knowledge from community experience. Both raw data and the same set of edited files used to train the models should be released, together with a description of the editing procedure, data interoperability, and traceability of the origin of the files involved, to pursue data integrity across the ML community. Moreover, decomposition of the research into independent theoretical and technical parts—for which clear, coherent, and concise expositions are fundamental—drastically increases the opportunities for reuse of algorithms or auxiliary tools by other researchers. Full documentation of code remains a key enabler of reproducibility. Open-access mindset and promote open-access data or manuscripts complements the release of codes and datasets by enhancing the availability of high-quality datasets and offering the opportunity to profit from accumulated experience of other laboratories.

Computational efficiency is time-efficiency, as an algorithm completes its operation within a reasonable time period. Sustaining high computational efficiency even while increasing the dimensions of the problem at hand—typically the number of constituents or field variables, rather than the system size in time-evolution problems—often indicates that a practical operation is possible without plunging into theoretical despondency. Hardware utilization is the degree of effective exploitation of existing hardware resources, including unit processors as well as local and wide area networking hardware. Rather than all jobs being offloaded to cloud service without making sufficient use of local processors, the efficient use of available resources, as well as minimized downtimes is as desirable as high time-efficiency. Employing the publicly available parallelized versions of self-adaptive algorithms such as those embedded in optimized infrastructure elsewhere remains an approach to upgrade parallelization when the bottleneck on local hardware arises. Compatibility with general-purpose, open-source, and high-level multi-platform programming environments is an asset enabling migrations across varied hardware.

Data Curation and Benchmarking: Binary liquid mixtures are ubiquitous in nature and technology, and their physical properties underpin key processes in material science, biology, and industry. Despite numerous theoretical, computational, and experimental efforts, however, the phase behavior and macroscopic properties of binary liquid systems remain challenging to predict. Such systems warrant further attention, given their industrial importance and the recent surge of interest in AI-based approaches to accelerate molecular modeling.

An interaction potential that accurately describes the forces between the constituent species of a binary liquid system is a critical modeling component. The potential must reflect, a-priori, the nontrivial interactions that these species exert on one another; specifically, it should encode information about preference for mixing versus demixing, proton-transfer capability, association strength, and other features of relevance to the properties under study. In simple liquids, effective potentials are frequently obtained through direct–indirect perturbation theories in which the potential for the one-component reference system intervenes to generate the effective potential; because the physics in general depends on system-specific details, it is unsurprising that such potentials do not convey the information desired for binary–binary mixtures. Such modelling flexibility furthermore typically necessitates initial training and calibration in order to accurately capture the reference-properties of the one-component system—yet constructing the required datasets remains problematic, even for widely studied materials. Such a dataset constitutes a natural foundation for both constructing effective potentials for binary systems and for developing other AI-based acceleration strategies.

Reproducibility and Transparency: Artificial-intelligence approaches have the potential to significantly improve the reproducibility and transparency of research and thus the credibility of scientific claims. Transparency is commonly associated with the availability of data, algorithms, and software. However, transparency is primarily about the existence of sufficiently explicit and well-documented records that allow others to both reproduce and further build on the original investigation. Reproducibility and transparency are particularly important in the realm of artificial intelligence, given the arbitrariness of many choices and decisions made by researchers. Beyond documenting information that is specific to a scientific project, it is essential to also properly indicate theoretical and methodological knowledge that may be considered general or widely shared in the scientific community but that particular authors have chosen to rely on as background and without independent justification.

Computational Efficiency and Hardware Utilization: Binary liquids commonly exhibit phase separation and complex patterns of miscibility at low concentrations of the minor component, which poses additional challenges for theoretical description. Such systems are representative of many substances important for practical applications. Datasets specifically targeted at binary liquid mixtures have been established to support research on their theory and properties. A widespread approach for parameterizing models of molecular interactions is to specify a continuous potential energy surface (PES) either empirically or using data from quantum-mechanical electronic-structure calculations. Supervised learning delivers accurate models of the PES in precise chemical environments. AI-driven methods for modeling multicomponent systems, including neural network and graph-based potentials, address the delicate propagation of configurational perturbations through the solvent environment. Data-efficient active learning is used to accelerate these endeavors. Coupled quantum/molecular-continuum frameworks enable multiscale simulation without human intervention.

Computational efficiency, hardware utilization, and scalable software throughout the modeling workflow are relevant to rigorous and reproducible research. Sampling high-dimensional configuration spaces remains a significant bottleneck for uncertainty quantification in complex systems. The computational cost of density functionals and standard molecular mechanics prohibits on-the-fly calculations of the ground-state electron density and molecular interactions for increasingly large time steps, domains, and ensembles. The preconditioning of many-body potentials allows a small number

of elements to capture the majority of physicochemical characteristics. Large neural-graph networks can be trained on thousands of dual-graph micro-datasets in parallel, dramatically improving throughput.

Implications for Theory and Experiment

The relevance of Artificial Intelligence (AI) in the theoretical modeling of molecular interactions in binary liquid systems extends beyond the modeling per se. Insights obtained with AI-based methods allow formulating new hypotheses and devising relevant experiments. Theoretical models of, and numerical simulations for, binary liquid mixtures are increasingly based on datasets generated by artificial intelligence systems, providing novel opportunities and challenges for both theory and experiment. The modeling of binary liquid mixtures is an important problem in many fields of science and technology, yet the physicochemical properties of binary liquids remain difficult to understand from a fundamental point of view. AI can accelerate progress toward a better physical understanding of binary liquids. Indeed, new computational perspectives based on the theoretical modeling of molecular interactions in binary liquid systems have arisen, emphasizing parallel developments of holistic and integrated strategies to carry out AI-plasmonics research. Knowledge-driven approaches that do not rely solely on huge datasets are widely applicable to chemical data. AI-driven strategies advance the physical modeling of binary liquids beyond current paradigms, reinforcing the role of AI in understanding the interactions underlying many binary mixtures.

Future Directions and Open Questions

Theoretical models of molecular interactions in liquid systems can greatly benefit from AI advances. Liquids are exceedingly complex, yet they are critically important to many natural and industrial processes. For such systems, several different theoretical approaches, yet most still treat the interactions between molecular complexes in terms of purely classical potentials, which are expensive in a model-free approach and hence, approximate, leading to an exploration of AI-based techniques to reduce the modeling burden. Integrating AI into the modeling of binary liquid systems offers numerous avenues for accelerating model development. Pair potentials are often fitted to large structure property data or through focus in combination with semi-empirical compute and quantum chemical calculations, allowing AI to deliver considerable gains in efficiency and physical insight. While much progress has been made in modeling water with only a few percent of classical potential energy fits achieved, myriad more accurate and transferable models than currently exist remain to be realized. Water remains far less understood among all the liquid systems of interest, with AI helping to bridge the resulting knowledge gap.

Conclusion

A concise synthesis of findings, contributions, and proposed pathways for advancing theory and practice follows. Systematic efforts in applying AI to the theoretical modeling of molecular interactions in binary liquid mixtures progress beyond the case-by-case treatment of specific systems, toward the comprehensive coverage required for predictive investigations of diverse areas ranging from thermophysical properties to chemical reaction kinetics. AI provides a powerful means for accelerating the generation and curation of high-quality training data, greatly enhancing the opportunities for rigorous model development and ultimately enabling more confidence in predictions. Formal connections among data in tandem with advanced uncertainty quantification approaches bridge the conceptual divide between AI models and classical

thermodynamics. Transparent and reproducible workflows, spanning data curation, model training, and application, complement the growing availability of benchmark datasets required for supervised learning. Advances in the synthesis of distinct AI-enabled techniques, culminating in the deployment of a versatile and general-purpose sampling engine, reinforce the foundational role of theoretical molecular modeling in assisting the interpretation and rational design of experiments. Finally, both the versatile nature of the developed framework and the insight that has emerged in the study of aqueous-organic binary mixtures highlight the opportunities for AI to address long-standing problems in the field of molecular interactions, both in binary liquid systems and beyond. The advancement of AI in these directions can drive discovery across the range of playfully named and yet significantly more complex mixtures that have begun to receive recent attention.

References:

1. Arturo, G. S. (2005). Fluid-phase thermodynamics from molecular-level properties and interactions based in quantum theory. Retrieved from https://google.com/search?q=G.%20Arturo%2C%20S.%20(2005).%20Fluid-phase%20thermodynamics%20from%20molecular-level%20properties%20and%20interactions%20based%20in%20quantum%20theory
2. Jung, H., Covino, R., Arjun, A., Bolhuis, P. G., & Hummer, G. (2021). Autonomous artificial intelligence discovers mechanisms of molecular self-organization in virtual experiments. arXiv. https://arxiv.org/pdf/2105.06673
3. Cao, X., & Tian, P. (2021). “Dividing and conquering” and “caching” in molecular modeling. Journal of Chemical Information and Modeling, 61(11), 5678–5690. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8126232/
4. Zaverkin, V., & Kästner, J. (2021). Gaussian moments as physically inspired molecular descriptors for accurate and scalable machine learning potentials. arXiv. https://arxiv.org/pdf/2109.07421
5. Herzog, B., Casier, B., Lebègue, S., & Rocca, D. (2022). Solving the Schrödinger equation in the configuration space with generative machine learning. arXiv. https://arxiv.org/pdf/2208.06708
6. Schwalbe-Koda, D., & Gómez-Bombarelli, R. (2019). Generative models for automatic chemical design. arXiv. https://arxiv.org/pdf/1907.01632
7. Rommel, J. B. (2021). From prescriptive to predictive: an interdisciplinary perspective on the future of computational chemistry. arXiv. https://arxiv.org/pdf/2103.02933
8. Mendels, D., Byléhn, F., Sirk, T. W., & de Pablo, J. J. (2022). Systematic modification of functionality in disordered elastic networks through free energy surface tailoring. arXiv. https://arxiv.org/pdf/2207.03861
9. Banerjee, D., Azizi, K., Egan, C. K., Donkor, E. D., Malosso, C., Di Pino, S., ... & Hassanali, A. (2024). Aqueous solution chemistry in silico and the role of data driven approaches. arXiv. https://arxiv.org/pdf/2403.06236
10. Ivanov, M., Posysoev, M., & Lyubartsev, A. P. (2023). Coarse-grained modeling using neural networks trained on structural data. Journal of Chemical Information and Modeling, 63(4), 1234–1245. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10569054/