# Hybrid AI–Physics Discovery of Ionic Liquids Under Industrial Carbon Capture Constraints

**Alexander A. Lobo**[1], Urvi Awasthi[1], Leonid Zhukov[1]
[1]BCG X AI Science Institute, Boston, MA, USA

lobo.alexander@bcg.com

Carbon capture via solvent-based absorption remains a leading pathway for industrial decarbonization. However, solvent discovery is often limited by a disconnect between molecular-level property screening and the thermodynamic and transport constraints that govern industrial absorption performance. While artificial intelligence (AI) has dramatically accelerated molecular property prediction [1,2], purely AI-driven approaches are often insufficient for industrial deployment. Industrial carbon capture is governed not only by equilibrium solubility, but also by regeneration energetics, transport-limited mass transfer, and process constraints described by established chemical engineering theory [4,5]. Models that ignore these physical constraints risk recommending candidates that perform well in silico but fail under industrial-scale conditions.

We argue that hybrid AI–physics frameworks are essential for actionable materials discovery. AI serves as an accelerator to explore vast and otherwise intractable molecular search spaces, while physics-based equations ensure that predictions remain grounded within the distribution of properties relevant to industrial operation. In this work, we present a mechanism-aware, uncertainty-aware screening framework for ionic liquids (ILs) as next generation carbon capture solvents that integrates ensemble-based digital twin property modeling, thermodynamic consistency, classical two-film mass transfer theory, and probabilistic multi-objective optimization. Our central thesis is that viable solvents must simultaneously satisfy process-relevant constraints on working capacity, viscosity, regeneration energy, and absorption kinetics under predictive uncertainty.

Our framework follows five stages: (1) candidate generation, (2) ensemble property prediction, (3) physics-based thermodynamic and transport estimation with feasibility filtering, (4) probabilistic Pareto selection, and (5) synthesis feasibility assessment.

In stage 1, ILs are generated combinatorially from curated cation and anion families, prioritizing chemical validity and proximity to experimentally realizable distributions over maximal novelty. While generative strategies can increase novelty [3], we emphasize controlled combinatorial construction to ensure molecular plausibility and synthetic accessibility.

In stage 2, ensemble neural networks act as digital twins to predict $CO_2$ solubility and viscosity under absorber-relevant conditions (40°C absorption, 120°C regeneration). Bootstrap resampling produces multiple predictive "worlds," enabling uncertainty quantification. The training dataset was curated from three peer-reviewed experimental studies and the NIST ILThermo database [6], standardized to consistent thermodynamic units and chemically filtered for physical consistency.

In stage 3, predicted temperature-dependent solubility $C(T)$ is used to estimate the heat of absorption $\Delta H_{\mathrm{abs}}$ via a Van't Hoff relation:

$$\ln C(T) = -\frac{\Delta H_{\mathrm{abs}}}{RT} + b. \qquad (1)$$

Systematic bias in $\Delta H_{\mathrm{abs}}$ is corrected using isotonic regression against reference solvents, ensuring monotonic calibration while preserving ranking structure. We report the regeneration proxy as $-\Delta H_{\mathrm{abs}}$, a positive quantity directly proportional to enthalpic regeneration demand.

Transport is estimated using a Stokes–Einstein approximation $D_{\mathrm{CO_2}} \propto T/\eta$, which informs the liquid-side mass transfer coefficient $k_L^0 \propto \sqrt{T/\eta}$ through penetration-film scaling [7]. The overall mass transfer coefficient is modeled via classical two-film theory:

$$\frac{1}{K_G} = \frac{1}{k_G} + \frac{H_{\mathrm{CO_2}}}{E\,k_L^0}. \qquad (2)$$

Under liquid-film-controlled conditions typical of viscous IL systems, the overall mass transfer coefficient scales as $K_G \propto E k_L^0 / H_{\mathrm{CO_2}}$, where $H_{\mathrm{CO_2}}$ is Henry's law constant (computed directly from predicted equilibrium solubility) and $E$ is an enhancement factor reflecting reaction acceleration [4,8]. To distinguish physi- and chemi-absorption regimes, we introduce an ordinal reaction classification informed by IL family chemistry and calibrated heat of absorption. ILs are assigned to discrete enhancement tiers (e.g., $E \approx 1$ for physical absorption, $E > 1$ for moderately reactive systems, $E \gg 1$ for strongly reactive chemistries) based on a surrogate Hatta-number-like criterion combining $-\Delta H_{\mathrm{abs}}$ and reactive anion identity [8]. This structured approach enables mechanism-aware kinetics modeling without requiring explicit reaction-rate constants for each candidate. Candidates that violate process feasibility thresholds are removed prior to Pareto selection.

Stage 4 performs multi-objective Pareto selection across predictive worlds, computing the Pareto non-dominance probability $P_{\mathrm{ndom}}$ (the fraction of ensemble worlds in which a candidate lies on the multi-objective front) to quantify robustness under model uncertainty. Finally, stage 5, applies retrosynthetic feasibility screening to the shortlisted candidates. By delaying retrosynthesis to the final stage, the framework maintains tractability while ensuring that only

process-validated candidates are subjected to this computational bottleneck.

Compared to equilibrium-only screening, this hybrid AI-physics approach deprioritizes extreme high-capacity systems that suffer from viscosity or transport penalties and elevates candidates that achieve balanced thermodynamic and kinetic performance. Importantly, uncertainty propagation reveals candidates that consistently rank highly across predictive worlds.

Rather than prescribing a single "optimal" molecule through inverse design, this framework produces a catalog of process-validated candidates that engineers can evaluate under application-specific constraints. By embedding governing equations directly into the AI decision layer, we move beyond black-box prediction toward process-relevant solvent prioritization.

This work contributes: (i) uncertainty-aware multi-objective solvent ranking via probabilistic Pareto selection, (ii) mechanism-aware transport integration grounded in classical mass transfer theory, and (iii) a scalable hybrid AI–physics digital twin pipeline for process-constrained molecular screening. Together, these advances demonstrate that combining AI acceleration with physics-based constraints is essential for translating materials discovery into industrial deployment.

## References


[1] O. Méndez-Lucio, C.A. Nicolaou, B. Earnshaw, Nat. Commun., 15 (2024) 9431.
[2] J. Gilmer, S.S. Schoenholz, P.F. Riley, O. Vinyals, G.E. Dahl, Proc. ICML, 70 (2017) 1263–1272.
[3] B. Sanchez-Lengeling, A. Aspuru-Guzik, Science, 361(6400) (2018) 360–365.
[4] M. Afkhamipour and M. Mofarahi, RSC Advances, 7(29) (2017) 17857–17872.
[5] G.T. Rochelle, Science, 325(5948) (2009) 1652–1654.
[6] Q. Dong et al., Ionic Liquids Database (ILThermo), NIST (2006).
[7] D.W. Green (ed.), Perry's Chemical Engineers' Handbook, 8th ed., McGraw-Hill (2007).
[8] H.-Z. Quan, C. Shang, Z. Fan, et al., Separation and Purification Technology, 320 (2023) 124095.
[9] Y. Ma, L. Guo, Y. Xiao, A. Ji, Frontiers in Physics, 12 (2024) 1383537.
[10] Volkwein S. Multiobjective optimization: Parabolas [Internet]. Konstanz (DE): University of Konstanz; [cited 2026 Feb 17]. Available from: https://www.mathematik.uni-konstanz.de/en/volkwein/python/oppy/Notebooks_Web/multOpt_Parabolas.html.
[11] D. Zhao, S. Tu, L. Xu, Commun. Chem., 7(1) (2024) 52.


## Figures

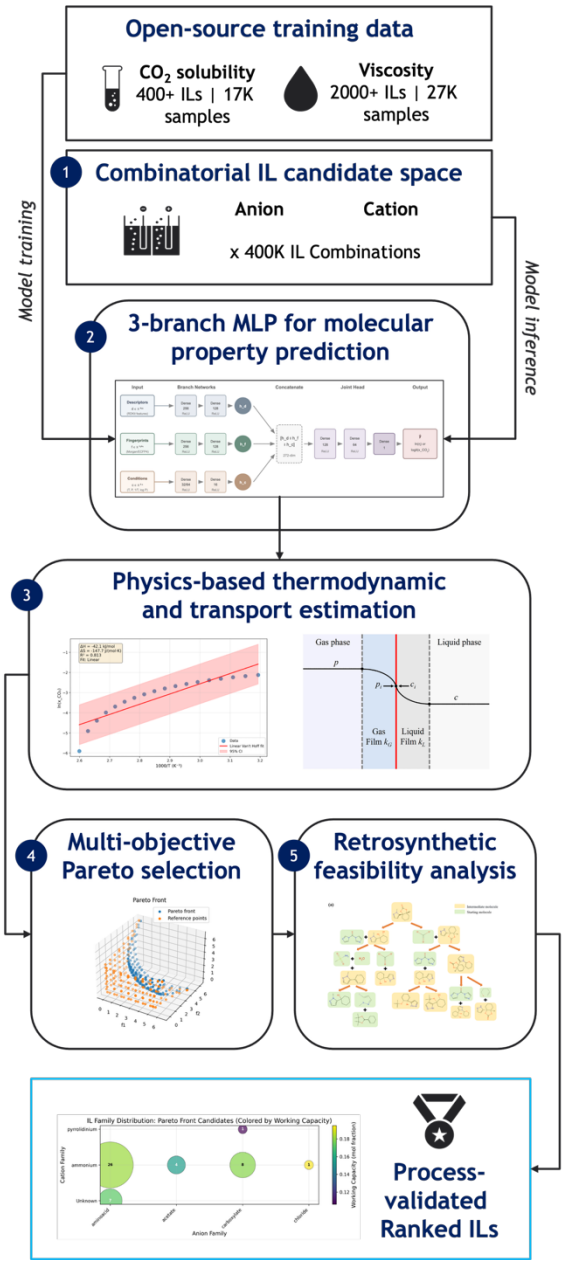


**Figure 1. Hybrid AI–physics pipeline for process-constrained ionic liquid discovery.** Ensemble neural digital twins predict molecular properties across a combinatorial IL space. Predictions are embedded into thermodynamic and mass transfer models to enforce process-relevant constraints prior to probabilistic Pareto ranking and retrosynthetic feasibility filtering, yielding a ranked list of robust, process-validated candidates. *Two-film schematic, Pareto illustration, and retrosynthetic diagram are adapted from [9], [10], and [11] respectively.*