Artificial Intelligence-Enabled Optimization of Battery-Grade Lithium Carbonate Production

S. Shayan Mousavi Masouleh [1, 2], Corey A. Sanz [3], Ryan P. Jansonius [3], Samuel Shi [4], Maria J. Gendron Romero [4], Jason E. Hein [3], Jason Hattrick-Simpers [1, *]

[1] Canmet MATERIALS, Natural Resources Canada, 183 Longwood Rd S, Hamilton, ON, Canada

[2] Department of Materials Science and Engineering, McMaster University, 1280 Main St W, Hamilton, ON, Canada

[3] Telescope Innovations, 301-2386 E Mall, Vancouver, BC, Canada

[4] Department of Materials Science and Engineering, University of Toronto, 184 College St, Toronto, ON, Canada

Corresponding Author

* jason.hattrick-simpers@nrcan-rncan.gc.ca

Supplementary Information:

1. Experimental procedures:

Table S1. This table presents data from high-throughput experiments utilized to develop the yield optimization model. The first column lists the batch number from which the data were obtained. Batch 0 contains historical data acquired through the traditional grid search method. Data from batches 1 and 2 were generated during active learning cycles. The acquisition policy that suggested each input combination is noted in the rightmost column. Total concentrations of C, N, and Li were calculated based on the concentration of these elements in the solution mix before crystallization occurred. This involves proper dilution calculations from the initial unmixed LiCl and NH_4OH/CO_2 solutions. Total C concentration was adjusted by sampling from the carbonation reactor at various times after initiating CO_2 bubbling. Total N and Li concentrations were altered by mixing the NH_4OH/CO_2 solution with the LiCl solution in different volume ratios, or by changing the initial Li concentration in the LiCl solution. Data in the table are rounded to two digits after the decimal for clarity.

# Batch	Initial aqueous concentrations (before crystallization, M)			Measured Li concentration at equilibrium (M)	%Yield	Acquisition Policy
	Total Carbon	Total Nitrogen	Total Lithium			
0	0.50	4.50	1.00	0.72	0.28	Grid Search
0	1.00	4.50	1.00	0.66	0.34	Grid Search
0	1.50	4.50	1.00	0.74	0.26	Grid Search
0	1.00	6.00	1.00	0.62	0.38	Grid Search
0	1.50	6.00	1.00	0.66	0.34	Grid Search
0	0.50	4.50	1.50	0.74	0.50	Grid Search
0	1.00	4.50	1.50	0.68	0.55	Grid Search
0	1.50	4.50	1.50	0.74	0.51	Grid Search
0	0.50	6.00	1.50	0.75	0.50	Grid Search
0	1.00	6.00	1.50	0.82	0.46	Grid Search
0	1.50	6.00	1.50	0.68	0.55	Grid Search
0	0.50	4.50	2.00	1.01	0.50	Grid Search
0	1.00	4.50	2.00	1.04	0.48	Grid Search
0	1.50	4.50	2.00	0.78	0.61	Grid Search

0	0.50	6.00	2.00	0.98	0.51	Grid Search
0	1.00	6.00	2.00	0.72	0.64	Grid Search
0	1.50	6.00	2.00	0.69	0.65	Grid Search
1	2.49	6.00	2.74	0.73	0.73	High Yield
1	1.98	6.00	2.50	0.71	0.72	High Yield
1	2.54	5.25	2.50	0.76	0.70	High Yield
1	2.17	5.75	2.02	0.78	0.61	High Yield
1	2.55	5.75	3.45	0.69	0.80	High Yield
1	2.12	6.00	3.45	0.67	0.81	High Yield
1	2.56	5.00	3.21	0.81	0.75	High Yield
1	1.61	6.00	3.21	0.78	0.76	High Yield
1	1.71	5.75	1.79	0.82	0.54	High Yield
1	1.36	5.75	2.50	0.78	0.69	High Yield
1	2.49	6.00	4.16	0.69	0.83	High Yield
1	1.68	4.75	2.74	0.79	0.71	High Yield
1	2.05	4.25	1.79	0.89	0.50	High Uncertainty
1	2.56	5.00	4.16	0.81	0.81	High Uncertainty
1	1.81	3.75	2.74	0.85	0.69	High Uncertainty
1	0.69	3.25	1.31	0.75	0.43	High Uncertainty
1	2.49	6.00	1.31	0.88	0.33	High Uncertainty
1	1.90	5.75	0.80	0.80	0.00	High Uncertainty
1	0.77	1.50	1.11	0.86	0.23	Random
1	1.24	3.00	2.02	0.88	0.56	Random
1	2.54	5.25	1.11	0.96	0.14	Random
1	1.57	3.25	1.11	0.82	0.26	Random
1	0.82	2.50	1.11	0.75	0.32	Random
1	0.37	6.00	0.43	0.43	0.00	Random
2	0.31	5.33	0.83	0.65	0.22	Random
2	0.92	2.17	2.00	0.90	0.55	Random
2	0.75	3.50	0.50	0.50	0.00	Random
2	1.00	3.83	2.33	0.85	0.64	High Yield
2	1.37	3.00	3.00	1.08	0.64	High Yield
2	1.64	3.83	3.50	0.98	0.72	High Yield
2	1.53	5.33	3.67	0.82	0.78	High Yield
2	1.92	4.50	3.67	0.83	0.77	High Yield
2	1.99	5.00	1.83	0.84	0.54	High Yield
*2	0.17	0.33	0.17	0.17	0.00	High Uncertainty

*2	0.17	1.17	0.17	0.17	0.00	High Uncertainty
*2	0.17	2.00	0.17	0.17	0.00	High Uncertainty
*2	0.17	2.83	0.17	0.17	0.00	High Uncertainty
*2	1.00	2.00	0.17	0.17	0.00	High Uncertainty
*2	0.17	3.67	0.17	0.17	0.00	High Uncertainty
*2	2.50	6.00	0.17	0.17	0.00	High Uncertainty
*2	1.33	2.83	0.17	0.17	0.00	High Uncertainty
*2	0.00	0.00	0.00	0.00	0.00	High Uncertainty
*2	2.50	5.00	0.17	0.17	0.00	High Uncertainty
*2	2.00	4.00	0.17	0.17	0.00	High Uncertainty
*2	1.17	4.83	0.33	0.33	0.00	Random
*2	1.00	5.67	0.17	0.17	0.00	Random
*2	0.33	4.50	0.17	0.17	0.00	Random

** The experiments with batch numbers *2 were suggested for experimentation by the ML model. However, due to the low concentration of initial lithium and the control experiments conducted by the experimentalist in this study, they were considered extremely diluted. As a result, the output was automatically assumed to be zero.*

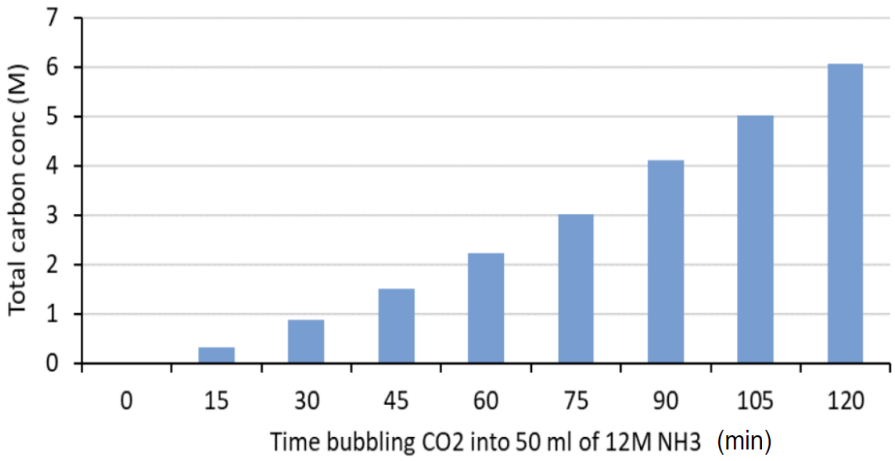

Figure S1: Total carbon concentration as determined by IR spectroscopy. Measured every 15 minutes by bubbling CO_2 into a 50 ml solution of 12M NH_4OH at a flow rate of 100 ml/min (temperature of reactor = 25 ºC).

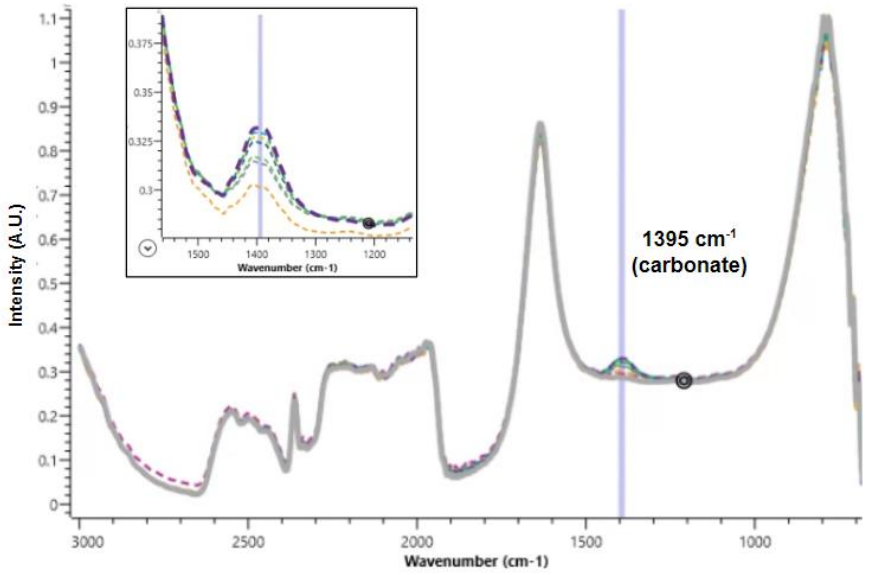

Figure S2: IR spectra (overlapping) used to build the graph shown in **Figure S1**. The intensity of the carbonate band at 1395 cm^{-1} was used to measure the total carbon concentration in the solution at each time interval. The inset is a close-up view on the carbonate band.

2. Influence of Temperature on Lithium Softening:

The impact of temperature on lithium softening in solutions with varying initial concentrations of carbon, nitrogen, and lithium was examined. Findings illustrated in **Figure S3** and detailed in **Table S2** reveal that variations in temperature below 66°C minimally affect lithium yields. **Table S2** presents the standard deviation and the percentage difference between the minimum and maximum final lithium concentrations at temperatures of 25°C, 48°C, and 66°C. These results indicate that temperature shifts within this range do not significantly alter the reaction outcomes. To accommodate slower reaction kinetics at lower temperatures, samples at 25°C and 48°C were maintained for seven days, compared to 24 hours at 66°C, following our experimental protocols.

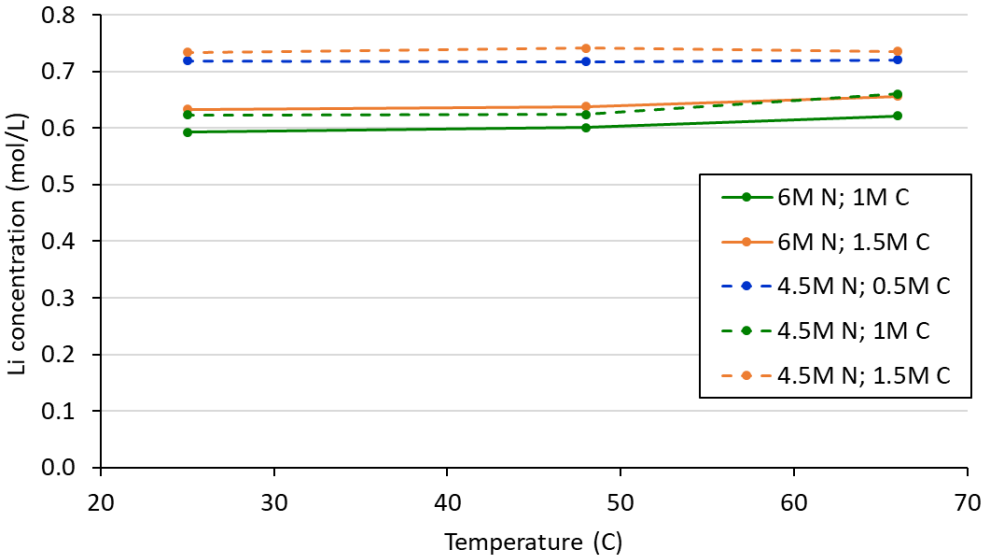

Figure S3: This figure demonstrates an example of lithium solubility at different temperatures across various nitrogen and carbon matrices, with an initial 1 M of lithium added to the solution. It shows the concentration of remaining lithium in the solution after equilibration. Each data point was collected from a solution prepared according to the procedure for 8 ml vial experiments outlined in the main manuscript. The only exceptions are that the samples at 25°C and 48°C were

mixed at these temperatures for 7 days instead of the standard 24 hours, to ensure full equilibration. All experiments were maintained at constant controlled temperatures during these periods (24 hours or 7 days). Thereafter, all experiments underwent a similar aliquot sampling and filtration process at room temperature for analysis. Given the short duration of this process, any potential temperature fluctuations are considered negligible and have not been factored into the results. Further experiments exploring the impact of temperature variations are detailed in **Table S2**.

Table S2: Temperature impact on final lithium concentration and yield. This table showcases the outcomes of experiments performed under various conditions at three distinct temperatures: 25°C, 48°C, and 66°C. It details the final lithium concentration in the solution (mol/L) and the yield (normalized to 1, solid) for each temperature setting. The table also presents the standard deviation and the percentage differences in final lithium concentrations observed across these temperatures under specific experimental conditions (Init C, Init N, and Init Li). For access to detailed raw data, refer to the files associated with batch 0 of the experiments available on our dedicated GitHub repository for this project.

Initial C (mol/L)	Initial N (mol/L)	Initial Li (mol/L)	Final Li (mol/L)			Li Yield			STD Final Li	% Difference Final Li
			25°C	48°C	66°C	25°C	48°C	66°C		
0.5	4.5	1.0	0.71	0.71	0.72	0.28	0.28	0.27	0.001	0.414
0.5	4.5	1.5	0.75	0.76	0.74	0.49	0.48	0.50	0.011	3.062
0.5	4.5	2.0	1.00	1.01	1.00	0.49	0.49	0.49	0.002	0.473
0.5	6.0	1.5	0.74	0.74	0.75	0.50	0.50	0.49	0.005	1.221
0.5	6.0	2.0	0.97	0.98	0.98	0.51	0.50	0.50	0.002	0.600
1.0	4.5	1.0	0.62	0.62	0.66	0.37	0.37	0.33	0.021	5.993
1.0	4.5	1.5	0.67	0.68	0.67	0.54	0.54	0.54	0.004	1.189
1.0	4.5	2.0	1.05	1.06	1.03	0.47	0.46	0.48	0.014	2.772
1.0	6.0	1.0	0.59	0.60	0.62	0.40	0.39	0.37	0.014	4.833
1.0	6.0	1.5	0.77	0.81	0.81	0.48	0.45	0.45	0.021	5.145
1.0	6.0	2.0	0.71	0.72	0.71	0.64	0.63	0.64	0.003	1.113
1.5	4.5	1.0	0.73	0.74	0.73	0.26	0.25	0.26	0.004	1.044
1.5	4.5	1.5	0.71	0.72	0.73	0.52	0.51	0.50	0.009	2.638
1.5	4.5	2.0	0.76	0.77	0.77	0.61	0.61	0.61	0.004	1.124
1.5	6.0	1.0	0.63	0.63	0.65	0.36	0.36	0.34	0.012	3.584
1.5	6.0	1.5	0.67	0.67	0.67	0.55	0.55	0.54	0.004	1.294
1.5	6.0	2.0	0.68	0.69	0.69	0.65	0.65	0.65	0.005	1.364

3. Experimentally viable ranges:

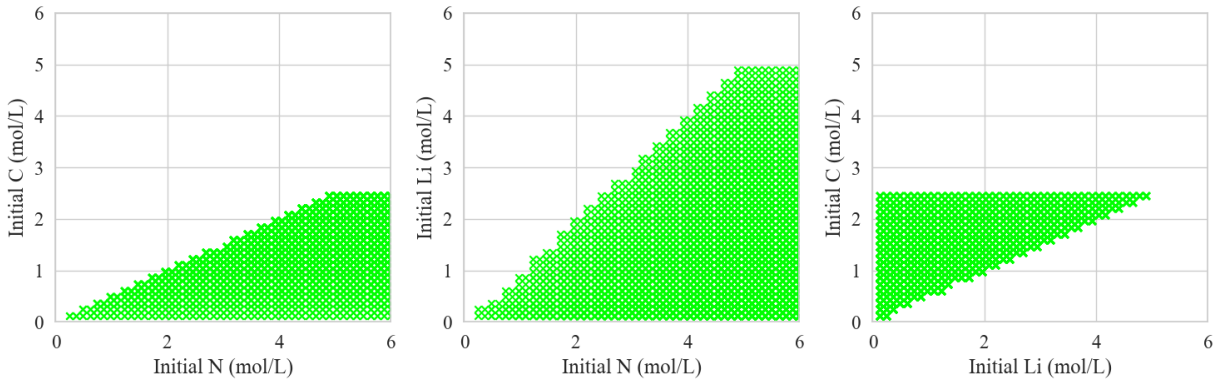

Figure S4: The green hashed area identifies the experimentally viable regions for initial concentration and permissible areas for data acquisition. Taking into account specific experimental constraints, such as stoichiometric ratios within the solution and varying solubility levels of different elements in water, not all initial concentrations within the explored chemical space were deemed practical. Further details governing the selection of combinations are provided in **Table S3.**

Table S3: Summary of the rules defining experimentally viable initial concentration combinations for nitrogen, carbon, and lithium in the chemical space. The constraints are categorized into solubility limitations and stoichiometry-controlled relationships, reflecting both the physical properties of the elements and the chemical interactions between them.

Rule Number	Element(s)	Constraint	Constraint Type
1	Nitrogen	Less than 6 mol/L	Solubility Limitation
2	Carbon	Less than 2.5 mol/L	Solubility Limitation
3	Lithium	Less than 6 mol/L	Solubility Limitation
4	Carbon, Nitrogen	Carbon $\leq 0.5 \times$ Nitrogen	Stoichiometry Controlled
5	Carbon, Lithium	Carbon $\geq 0.5 \times$ Lithium	Stoichiometry Controlled

Our data acquisition strategy segregates the data pool into three tiers, allocated proportionally at 50%, 25%, and 25%. Each tier samples data based on specific objectives, ensuring that no two sampled points in the chemical space are closer than a Euclidean distance of 1 mol/L. The three acquisition objectives or tiers considered are as follows: (1) lithium carbonate yield of the HTE experiments, where the aim is to acquire initial conditions with the highest predicted lithium yield; (2) uncertainty of the GPR predictions, where the acquisition policy focuses on points with the highest prediction uncertainty; and (3) random exploration of the parameter space, involving random selection of initial condition combinations.

Therefore, beginning with an available initial batch of experimental data (Batch 0), we constructed a preliminary GPR prediction of the chemical space. We conducted two rounds of active learning, with each round proposing 24 experiment suggestions. The initial round (Batch 1) primarily

targeted high lithium yield areas, with the majority of experiments sourced from high-yield regions. In contrast, the second round (Batch 2) emphasized broader space exploration based on the GPR model's uncertainty of the predictions (standard deviation). The specific acquisition breakdown for each round is detailed in **Table S4**.

Table S4: Distribution of each data batch according to different acquisition policies.

	Acquisition Policy			Total
	Yield	Uncertainty	Random	
Batch 0				17
Batch 1	12	6	6	24
Batch 2	6	12	6	24

4. AI-Enhanced Yield Optimization in Li-Brine Softening

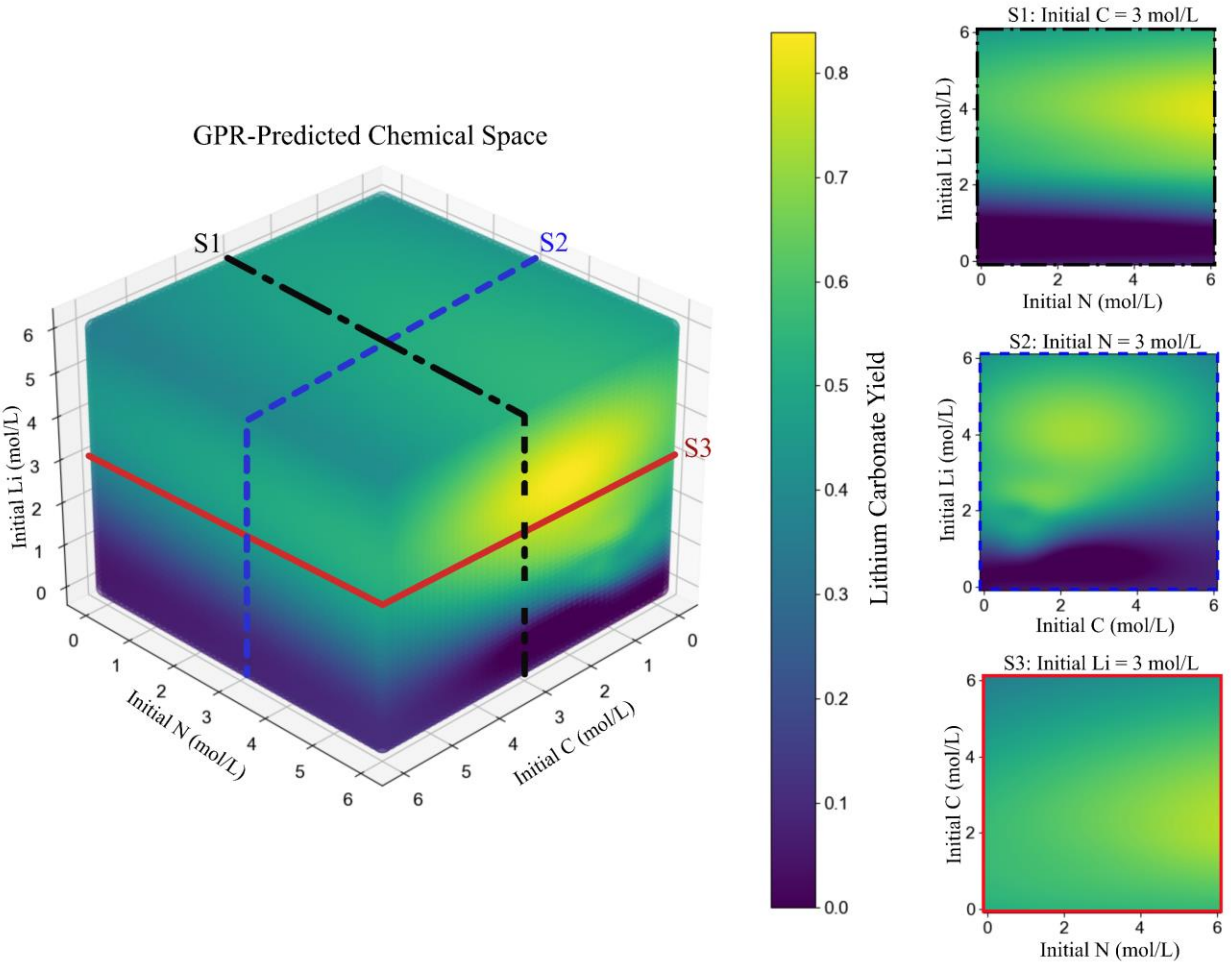

Figure S5: Three-dimensional visualization of the predicted chemical space using the GPR model, trained on all collected data (Batches 0, 1, and 2). The space is intentionally under-sampled to highlight the meshed nature of the chemical space. Three specific slices (S1, S2, and S3) are extracted and displayed on the right, providing detailed views of the predicted lithium carbonate yield landscape at different planes within the space.

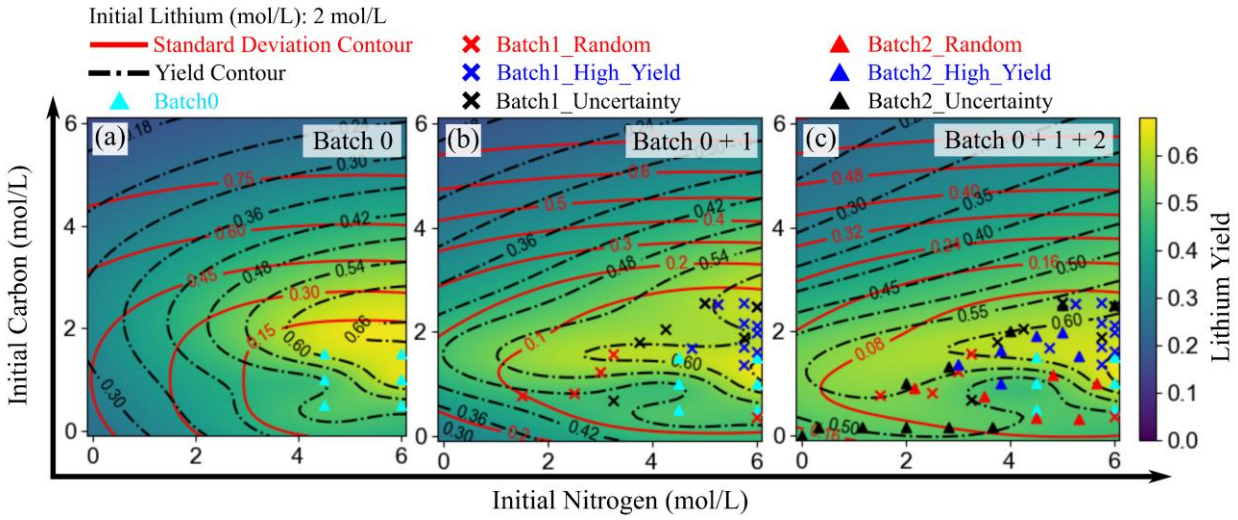

Figure S6: Visualization of Lithium Yield Predictions Using Gaussian Process Regression (GPR) in the Context of the Lithium Softening Process. The model was trained on data from three stages: (a) Batch 0 alone, (b) a combination of Batches 0 and 1, and (c) an aggregate of Batches 0, 1, and 2. These stages correspond to iterations 0 to 2 in the active learning cycle. The data points acquired at each stage are overlaid on their respective colormaps. Each colormap captures the lithium yield within a specific chemical space, holding the initial lithium concentration constant at 2 mol/L while allowing for variations in carbon and nitrogen levels between 0 to 6 mol/L. Black and red contours signify lithium yield and the associated GPR prediction uncertainty (measured as standard deviation), respectively.

5. Performance Analysis of Active Learning Strategies

During our active learning iterations, we gauged the learning quality by observing the uncertainty in the GPR model's predictions. As shown in Figure 3, uncertainty, measured by the standard deviation of the predictions, reduces with each iteration. Consequently, the areas with low uncertainty (or high prediction certainty) broaden. Using violin plots, we traced the prediction uncertainty throughout each active learning iteration. Additionally, we compared the uncertainty distribution from our tiered data acquisition with that from a fully random acquisition strategy, as illustrated in **Figures S7**.

The datasets for fully random acquisition, encompassing both initial parameters and predictions, were created using the surrogate GPR model trained on all the available experimental data. To mimic real-world experimental conditions, we introduced a random fluctuation of up to 10% into the model's input parameters, specifically the initial concentrations, as well as the predicted yields. The fully random acquisition distributions shown in **Figures S7**, represent the average standard deviations of predictions taken over 100 different random seeds for each iteration. Breaking it down further, for iteration 1, we amalgamated the data from batch 0 (consisting of 17 experiments) with data randomly sampled to match the size of batch 1 (24 experiments). This process was repeated across 100 distinct random seeds. Similarly, for the subsequent iteration, data from batch 0 was combined with data randomly drawn to cumulatively match the sizes of batches 1 and 2, amounting to 48 experiments in total. This amalgamation was also repeated across 100 different random seeds to guarantee the representativeness of the acquired datasets.

Figure S7 highlights that the active learning training process effectively reduced the GPR model's uncertainty. This reduction in prediction uncertainty, coupled with the concentration of uncertainties around the mean in the violin plot, suggests an enhanced robustness of the model. Notably, even as the second iteration targeted high-uncertainty areas from the initial predictions, the model remained consistent when exposed to new data. This consistent performance is indicative of the robustness and predictive fidelity of our active learning methodology.

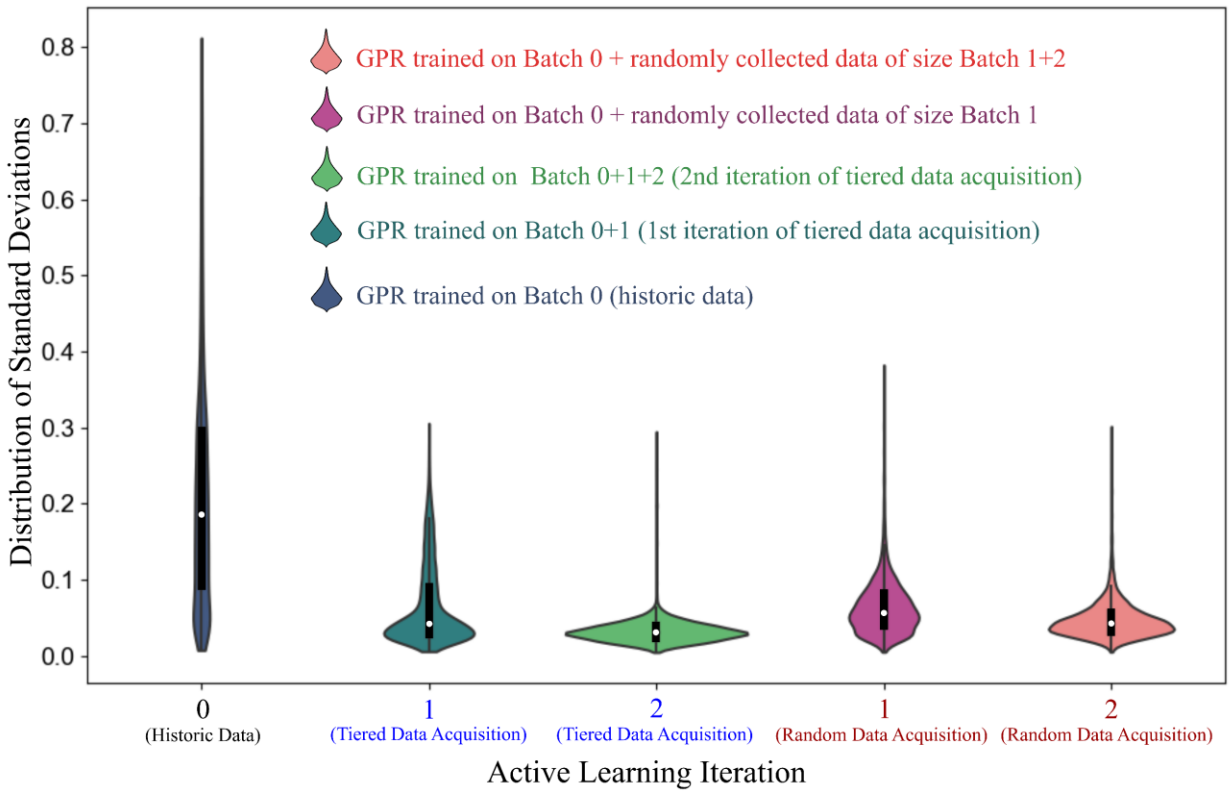

Figure S7: Violin plots depicting the overall standard deviation in Gaussian Process Regression (GPR) models across various active learning iterations. The plots compare the performance of models trained using the tiered data acquisition strategy employed in this study with those using fully random data acquisition.

6. Evaluation of Data Acquisition Policies

To gauge the influence of each data acquisition tier (high yield, high uncertainty, random) used in this study, we evaluated the individual contributions of each tier. To do this, we trained models using the base data (Batch 0) combined with data sourced through each specific policy per iteration. We used two metrics to quantify the contribution of each acquisition method: mean square error (MSE) and information gain, as depicted in **Figure S8**.

The MSE metric contrasts a surrogate GPR model—trained using all available data from Batches 0, 1, and 2—against models trained using the base data (Batch 0) combined with data from each specific data acquisition tier (policy). **Figure S8** shows that focusing on high-yield areas substantially reduces the MSE of the model, enhancing the accuracy of high-value predictions. Moreover, targeting high uncertainty areas considerably lowers the MSE by addressing regions where the model's predictive certainty is low. In contrast, random data acquisition has a less pronounced effect on model performance when compared to the other two data tiers.

Information gain, a measure derived from information theory, determines the dataset's entropy in relation to a benchmark, which in our case is the entropy of predictions using only the base data (Batch 0). We computed entropy with a continuous differential entropy formula, taking the standard deviation of the GPR predictions to represent local entropy. As displayed in **Figure S8**, both high yield and uncertainty-focused acquisition strategies provide similar information gains.

It's worth highlighting that although random data acquisition might not have a substantial impact compared to the other methods, excluding it could lead to biased data collection. Even without observing unexpected outcomes in our predictions and chemical space experiments, random data sampling could potentially identify regions prone to such anomalies. Hence, we deem all three data acquisition tiers crucial for this study.

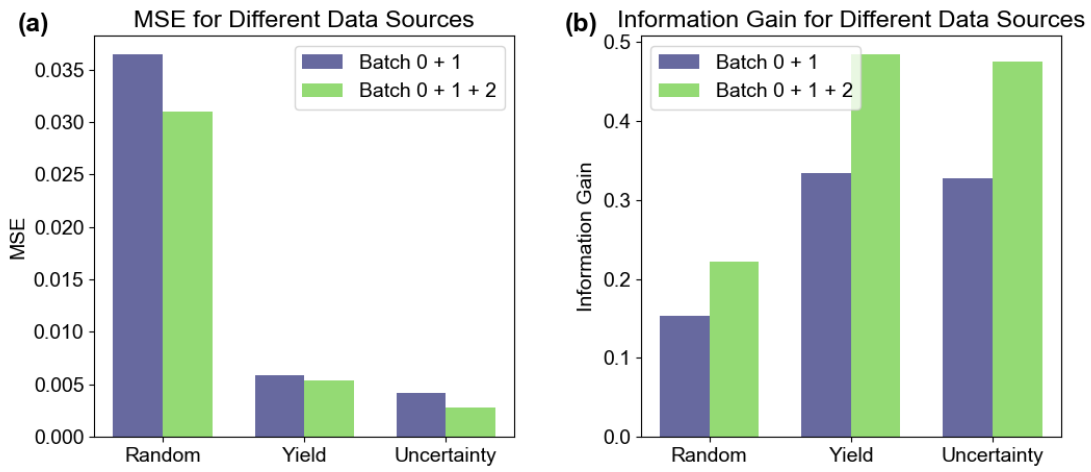

Figure S8: Assessment of the performance of various data sources in the active learning process (random acquisition, high yield acquisition, high uncertainty acquisition) using: (a) the mean squared error comparison of models trained with different data sources (acquisition policies) against those trained with all available data; (b) information gain for different data sources (acquisition policies) versus the model trained using only initial data (Batch 0).

7. Comparative Analysis of Machine Learning Models

While the primary aim of our study is not to demonstrate the superiority of GPR over other models, we conducted a comparative analysis to evaluate its performance relative to commonly used regression models such as Random Forest Regressor (RFR), XGBoost (XGB), and Support Vector Machine Regression (SVR). Each model was selected based on its ability to handle non-linear relationships and complex patterns: GPR with a Matern kernel for its smoothness and flexibility, RFR for robust non-linear regression, XGB for gradient boosting efficiency, and SVR for its kernel-based handling of non-linearities.

We conducted 5-fold cross-validation for each active learning iteration, repeating this process 10 times with different random seeds to ensure statistical robustness. The outcomes, as depicted in **Figure S9** using the mean squared error (MSE) metric and in **Figure S10** using the R2-score, indicate that GPR generally outperformed other models, particularly during the first iteration when fewer data points were available for training. This phase (first active learning iteration) demands heightened extrapolative capabilities, which are critical in active learning settings where models must predict beyond the observed data range. Tree-based methods such as RFR and XGB showed some limitations for extrapolation, which is expected given their design. However, as shown in the results from the second iteration (**Figure S10**), these methods performed better within the scope of observed data as sample sizes increased, although concerns about their extrapolation capabilities persist. Conversely, SVR consistently demonstrated inferior performance compared to the other models. It is important to note that this analysis specifically reflects the performance of these models on the data used in this study and is not intended to set general performance benchmarks.

These results highlight the suitability of GPR for our study, particularly due to its superior extrapolation abilities and robust uncertainty quantification, which significantly enhance the active learning process.

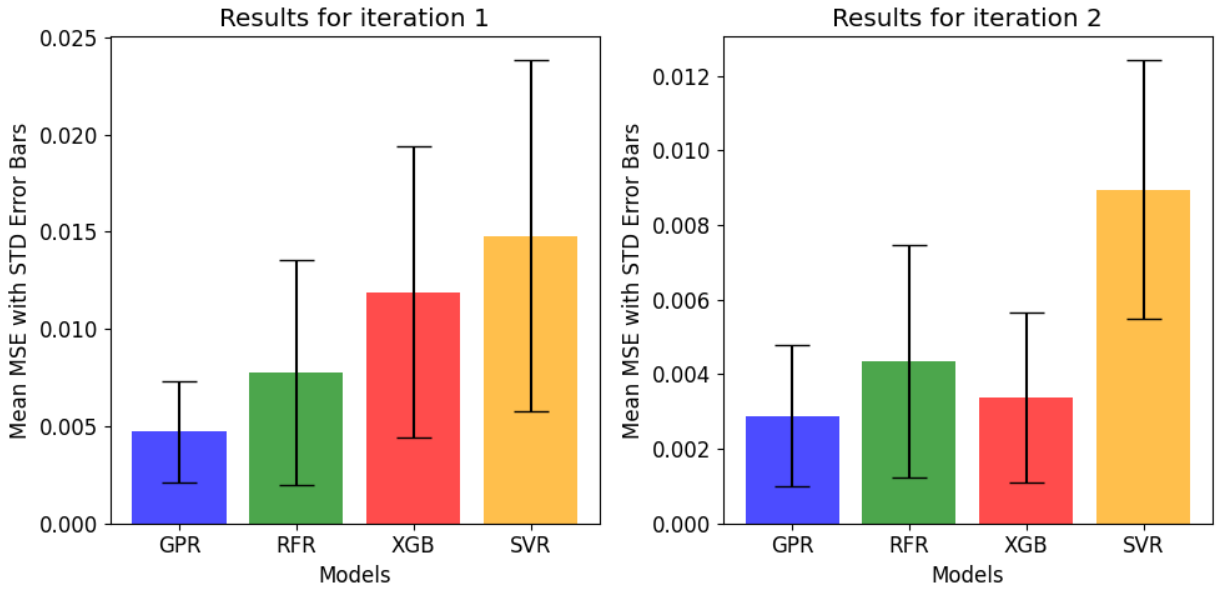

Figure S9: Performance of different models using mean squared error (MSE) metric. This figure illustrates the MSE across various models during the first and second active learning iterations, with the plot on the left depicting the first iteration and the plot on the right showing the second iteration.

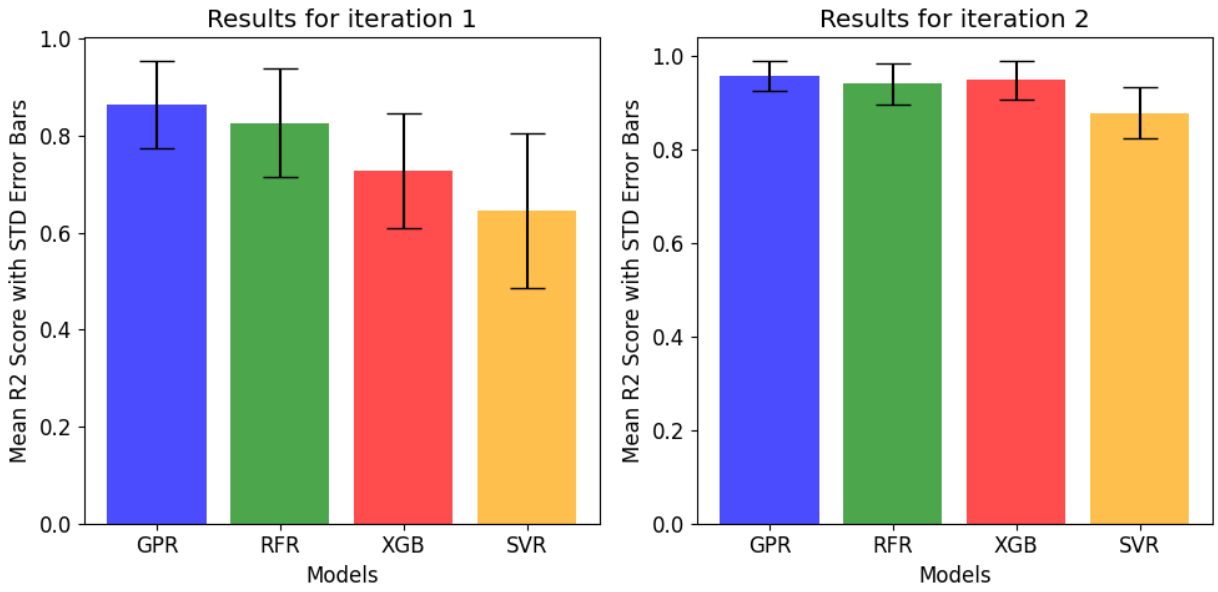

Figure S10: Model performance comparison using R2-score. This figure illustrates the R2-scores for various models during the first and second active learning iterations, with the plot on the left depicting the first iteration and the plot on the right showing the second iteration.