GEMS - Guided Evolutionary Molecule Design for Sustainable Chemicals

Coelina Robinson Franziska Weissbach Kjell Jorner Mennatallah El-Assady Christina Humer

ETH Zürich, Zurich, Switzerland

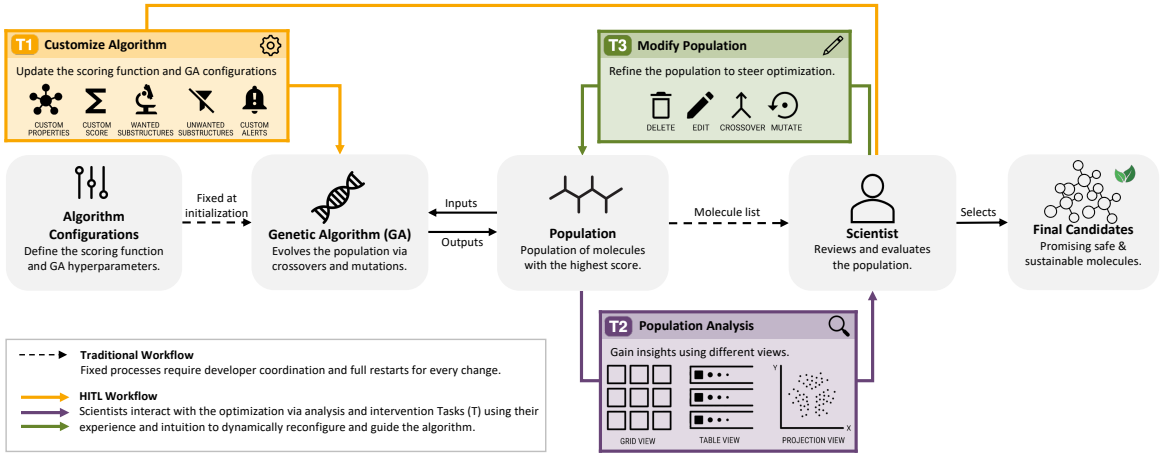


Figure 1: A workflow comparison of traditional genetic algorithm pipelines *vs.* our proposed Human-in-the-Loop (HITL) framework for sustainable molecule design. Unlike traditional linear, automated processes, our interactive workflow allows scientists to steer optimization, inspect candidates, and intervene in the evolutionary loop in real time.

Abstract

Designing safe and sustainable chemicals is critical to combat chemical pollution in our environment. Computational and AI-assisted methods have been developed to aid de novo molecule design. However, data on the environmental impacts of chemical compounds are sparse, resulting in low-fidelity machine learning (ML) oracles and unreliable candidate proposals. Furthermore, many automated molecular design approaches rely on numerical scoring functions that cannot fully capture the nuanced chemical intuition of expert scientists required for real-world molecular design. Instead, we present GEMS—an interactive visual analytics tool for human-in-the-loop molecular optimization that lets domain experts directly collaborate with an evolutionary genetic algorithm. Users continuously guide the search using domain knowledge through high-level, parametric modification of the scoring function alongside direct, granular control over molecule populations. GEMS requires no programming or expertise in ML or evolutionary optimization. A usage scenario demonstrates its application in designing sustainable antioxidant alternatives, and interviews with domain scientists provide feedback on its usefulness.

Index Terms: Human-in-the-loop optimization, molecular design, genetic algorithms, green chemistry, safe-and-sustainable-by-design, interactive visualization.

1 Introduction

Chemical pollution poses a critical threat to human health and our Earth's ecosystems. Rapid chemical production has outpaced our ability to assess compound safety, necessitating a shift from reactive testing to proactive, safe-and-sustainable-by-design (SSbD) methodologies [7]. Designing alternative molecules that retain the desirable functional properties of existing pollutants, such as PFAS, antioxidants, or refrigerants [4, 9, 26], while being safer and more sustainable, is therefore an important yet challenging task. Environmental safety and sustainability are difficult to predict due to limited data availability [32], and their assessment still relies on scientists' expertise and intuition. Such expertise spans synthetic feasibility and greenness, persistence estimation [22], and recognition of instability and reactivity, such as transformation products and interactions with organisms [15].

Existing computational and AI-assisted molecular design methods aim to support this process, but are often not easily accessible to domain scientists [29] because they require programming knowledge or developer support to use and adapt them to scientists' needs. Furthermore, current models remain limited by the accuracy of available oracles and the lack of any oracles for many relevant and often context-dependent properties [14]. As a result, effective molecular design still requires the supervision of domain experts to guide and refine the process [11, 21].

To address this gap, we developed an interactive visual tool that enables human experts to integrate their intuition for the relationship between molecular structure and diverse properties with a generative system for molecular design without requiring technical knowledge of ML models or evolutionary optimization methods. The platform allows users to explore and analyze generated molecules, steer the genetic algorithm, and iteratively refine results throughout the evolutionary process. Our approach combines efficient computational optimization with the generalizable, intuitive knowledge of human experts in an interactive workflow.

2 Background and Related Work

We provide an overview of genetic algorithms and methods for interactive molecular visualization and exploration, human-guided optimization, and their application to environmental chemistry.

E-mail: {crobinson, fweissbach, kjorner, melassad, humerc}@ethz.ch

2.1 Genetic Algorithms

A genetic algorithm (GA) is an evolutionary optimization method inspired by natural selection. GAs iteratively evolve a population of candidates [19]. In each iteration, the fittest candidates (*i.e.*, highest scoring according to a scoring function) are selected and modified by crossover and mutation operations. Crossover recombines molecular substructures of two parent candidates to produce novel offspring that inherit features from both parents. Mutation introduces local structural changes to individual candidates, exploring the search space. These operations can be handled stochastically or by integrating other models, such as LLMs [31].

GAs have proven to be strong baselines for molecular generation, often outperforming more complex generative models on standard benchmarks [28]. GEMS builds on graph-based GA operators that perform crossover and mutation directly on molecular graphs [3, 16].

2.2 Interactive Molecular Visualization and Exploration

Several visual analytics systems have been developed to support chemists in exploring and interpreting molecular data. A common approach is to use dimensionality reduction techniques to efficiently process the high-dimensional space of molecular features, enabling visual assessment of structural diversity and clustering [6, 12, 13]. To support systematic comparison and filtering across molecular properties, multi-attribute table views are a popular choice [10, 17]. Beyond exploration, recent systems have incorporated model explanations [12, 13, 23] and uncertainty visualization [13] to make data-driven predictions more transparent, and Menke et al. [18] have introduced a GUI to collect expert feedback for generative chemistry models.

Our work takes inspiration from existing research in the field and specifically CIME [12] in using a combination of structure, table, and projection views, but differs in a key aspect: rather than analyzing precomputed results or existing datasets, our tool enables users to directly steer a generative optimization process through iterative interaction with a genetic algorithm.

2.3 Human-Guided Molecular Optimization

Machine learning has become a central tool for molecular design, with approaches ranging from variational autoencoders and reinforcement learning to diffusion models to generate novel candidate structures [1]. However, current models still struggle to capture the nuanced, experience-driven judgments that chemists bring to molecular evaluation, such as synthetic feasibility or functional suitability, which are difficult to encode in a computational scoring function [5, 14] and necessitate trial-and-error adjustments [8]. This has motivated efforts toward human-in-the-loop generative modeling [20], where human experts guide model behavior via techniques like preference learning [5, 18, 27] or active learning [21]. Existing steering approaches, such as interactive visualizations of optimization campaigns in chemistry [13], typically mediate human input through a model—expert feedback updates a scoring or surrogate function, which then only indirectly drives the optimization.

Our tool takes a different approach: instead of channeling expert knowledge through a scoring function, it provides chemists with direct, continuous control over a GA acting as an evolutionary optimizer, transforming the expert from a queried oracle into an active collaborator who shapes the search at every iteration.

2.4 Safe and Sustainable Chemical Design

While human-in-the-loop approaches have begun to appear in drug discovery [18, 21, 27], our tool focuses on sustainability-oriented design. The benefit of interactivity is particularly important in SSbD because computational oracles for toxicity and persistence are less mature than for, *e.g.*, binding affinity, which is central to medicinal chemistry, and it is important to quickly identify and remedy reward hacking. The first generative modeling approaches to designing environmentally friendly chemicals have made progress [30, 33], but they remain constrained by property predictors trained on sparse data that often fail to generalize to novel structures and have no options to influence the generation process.

3 Goals and Core Tasks

The main goal of this tool is to support scientists in collaborating with a GA to design new, safe, and sustainable chemicals. It enables users to fine-tune the system to their specific use case without support from developers and to integrate domain knowledge and expert intuition to mitigate ML drawbacks. Based on discussions with domain scientists, we identified the following core tasks:

T1 Customize Algorithm

Scientists need to be able to run a genetic algorithm for molecular design out of the box, while retaining the ability to customize it for different research goals.

- T1.1 **Configuring key parameters** such as population size and number of generation cycles without the need for programming or computational modeling knowledge.
- T1.2 **Customizing the scoring function** and updating it throughout the evolutionary process to reflect specific research goals and changing knowledge on how to reach these goals.
- T1.3 **Defining structural constraints** to incorporate expert knowledge on desirable or undesirable substructures in the search.

T2 Population Analysis

Scientists need to analyze the evolving population to identify promising design directions and diagnose potential issues in the optimization process. Based on insights from this analysis, scientists can make informed decisions on how to adjust the scoring function or the population.

- T2.1 **Investigating candidates** in detail to help experts learn and adapt their knowledge of the search domain. Users should be able to investigate candidate features, structures, and scores and compare them to other candidates.
- T2.2 **Exploring the population** to assess whether the genetic algorithm is appropriately balancing exploration and exploitation of the search space.

T3 Modify Population

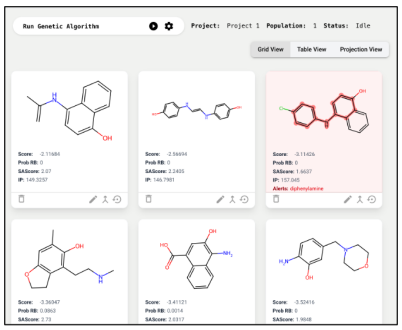
Scientists need to be able to manually adapt the population and modify candidates. This is in addition to the higher-level configuration of the scoring function and allows for a fine-grained way of steering the optimization process by utilizing users' internalized expertise.

- T3.1 **Removing molecules** from the current population allows users to exclude undesirable or implausible candidates as input for the next GA generation.
- T3.2 **Guiding mutation and crossover** operations to allow for a custom treatment of particularly promising candidates.
- T3.3 **Editing structures** of molecules to enable fine-grained control of the candidates beyond the generations of the GA.

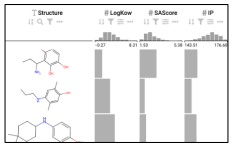
4 System Design

This section outlines the platform's workflow and details how individual components assist users in performing their specific tasks[1].

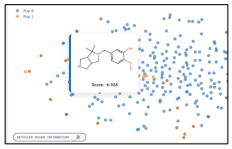
[1]Code available at: https://github.com/ETH-IVIA-Lab/gems.



(a) Grid View.



(b) Table View.



(c) Projection View.

Figure 2: The tool provides three views for exploration: (a) the Grid View, (b) the Table View, and (c) the Projection View.

4.1 Overview & Workflow

Figure 1 illustrates a traditional genetic algorithm workflow compared to the human-in-the-loop workflow enabled by GEMS. In the traditional setting, a developer programs the scoring function and configures the algorithm, after which the GA runs iteratively and returns the resulting population to the scientist, leaving little room for expert input or adjustment during the process. Human feedback is limited to inspecting the final output and adjusting the algorithm's initial configuration accordingly.

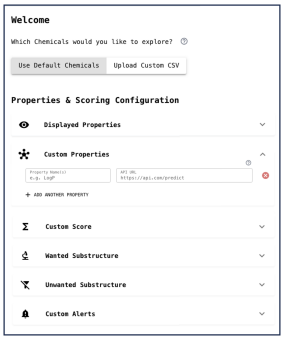
With GEMS, scientists can define and modify the scoring function themselves through an intuitive interface without requiring developer support. Crucially, both the scoring function and the algorithm's hyperparameters can be updated at any point during the process without restarting, allowing the search to be continuously steered between generations as new insights emerge (T1). The resulting molecules are visualized through three complementary views, a grid view, a table view, and a projection view (Figure 2), each supporting different analytical tasks (T2). Through these interfaces, users can inspect and analyze the population, intervene directly by deleting undesirable candidates, mutating or crossing over promising molecules, or manually editing structures (T3). Users can then re-run the GA on the modified population, closing the human-in-the-loop cycle and iteratively guiding the search toward desired molecular properties. This functionality is delivered via a lightweight web application to eliminate installation barriers, ensuring domain scientists can access the system regardless of their local computational environment.

4.2 Algorithm Configurations

To satisfy T1.1, the GA supports a single-click launch with sensible defaults, while still allowing advanced users to manually tune hyperparameters like population size. When launched, users are presented with an overview of configurable components, and the option to either upload their own dataset of molecules or use a predefined sample from the PubChem compound database (Figure 3a). Users are provided with a graphical scoring editor (Figure 3b) that lets scientists define and update scoring functions without programming knowledge or developer support. Users are presented with a default scoring function that balances functionality, synthesizability, and environmental impact, which can serve as a starting point or be fully reconfigured by adjusting property weights or adding or removing terms entirely (T1.2).

To give users full flexibility in scoring and allow them to include trusted external tools, external APIs can be integrated into the scoring function, extending it with custom molecular properties beyond the defaults (Figure 3a). Substructure-based rewards and penalties address T1.3 by allowing experts to encode structural intuition directly, for example, penalizing known persistent pollutant scaffolds without translating them into explicit scoring terms. Alerts can also visually flag substructures in the population view (Figure 2a), keeping expert knowledge visible throughout the process rather than only at configuration time.

Critically, none of these configurations are locked after initialization. Users can analyze the population (T2) throughout the process to update the scoring function, weights, and hyperparameters. This is a deliberate change from traditional GA pipelines, where the scoring function is fixed before execution. Here, real-time reconfiguration allows the search trajectory to respond to users' observations as the population evolves.



(a)

(b)

Figure 3: The configuration interface (a) allows users to upload a dataset, customize attributes and the scoring function. The expanded custom property tab shows API integration. The score editor (b) allows viewing and modifying the default scoring function.

4.3 Exploration and Analysis Views

The tool provides three distinct interfaces (Figure 2). The Grid View enables rapid visual screening. By embedding alerts directly into the 2D molecular drawings, chemists can instantly spot and remove undesirable candidates before running the next GA cycle. The Table View allows users to analyze, sort, and filter candidates across multiple properties. This design is valuable for SSbD evaluation, where chemists must balance conflicting objectives like maximizing chemical activity while minimizing environmental persistence, quickly filtering out candidates that achieve a high score but fail a single, critical safety threshold (T2.1). Finally, the Projection View addresses population-level structure. Embedding molecules into 2D space using Morgan fingerprints [24] allows users to identify search convergence or locate structurally similar neighbors to a promising candidate to better understand the effect of specific structural changes (T2.2). Together, these views enable a multi-faceted understanding of complex search spaces (T2) and support decision-making.

4.4 Interactive Steering and Refinement

Effective molecular design requires intervention at different levels of granularity depending on what the user observes. Deletion enables fine-grained filtering of undesirable candidates (T3.1), while manual mutation and crossover (Figure 4a) allow users to selectively apply evolutionary operators to promising candidates rather than leaving the recombination entirely to the GA (T3.2). For precise structural changes, an integrated drawing tool [2] (Figure 4b) enables edits beyond autonomous GA capabilities (T3.3). Users can also prompt an LLM in natural language to mutate or cross over molecules, which is useful when users have an abstract intuition that can be translated into various concrete structures. After any combination of interventions, the GA can be run on the modified population, closing the human-in-the-loop cycle.

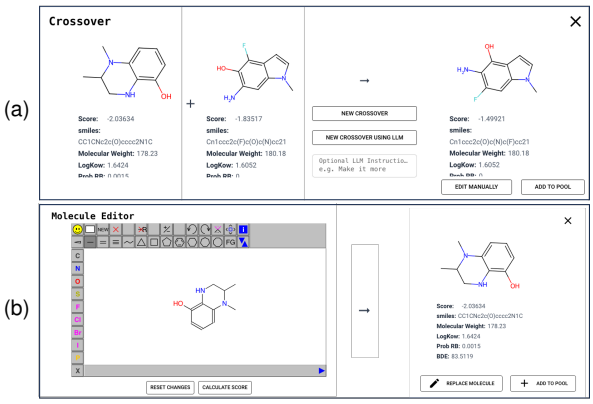


Figure 4: Interaction interfaces for manual population steering: (a) crossover and (b) direct structure editing.

5 Usage Scenario

To illustrate how our tool[2] can be used by scientists, we walk through the following scenario: Synthetic antioxidants are widely used in consumer products such as food packaging, skincare, and plastics. However, many have been flagged for environmental persistence and ecotoxicity, creating an urgent need for safer alternatives [26]. We illustrate this challenge through a scenario in which a chemist, Alice, uses our tool to identify novel antioxidant candidates that are both effective and environmentally friendly.

Alice begins by uploading a curated dataset of diverse phenolic compounds, leveraging the fact that phenol groups are the key component that makes many natural and synthetic antioxidants effective [26]. Her curated starting population ensures structural diversity while anchoring the search in a chemically meaningful space. She then configures the scoring function to balance four objectives equally: (1) antioxidant activity, (2) toxicity, (3) environmental persistence, and (4) synthesizability (Figure 3b). This reflects the multi-objective nature of safe-by-design molecular discovery. Drawing on domain knowledge, she additionally specifies structure-based penalties for tert-butyl substituents, which are common in commercial antioxidants but associated with environmental persistence [26], as well as for bisphenol A and estrogen-like scaffolds that are linked to endocrine disruption [25], meaning hormone interference. Furthermore, she creates an alert for diphenylamines, a motif that can degrade into toxic primary aromatic amines [26], allowing case-by-case judgment rather than penalty.

After saving her settings, Alice inspects the initial population in the Grid View, sorted by the scoring function, which shows each molecule's properties and score (Figure 2a). She removes molecules flagged with red alerts, and then switches to the Table View (Figure 2b) to sort by individual properties and delete candidates that perform poorly on any single criterion, ensuring a clean starting population. She runs the GA and observes which structural motifs persist across generations. She notices that halogenated phenol derivatives consistently exhibit high antioxidant activity, and the penalty for their low biodegradability is insufficient to outweigh this advantage. In response, she increases the weight of biodegradability in the scoring function and uses the LLM-guided mutation feature to suggest alternative substituents for a selected candidate. Afterwards, she runs more GA cycles.

Lastly, she uses the Projection View (Figure 2c) to assess the diversity of the evolved population. While later generations occupy a higher-scoring but increasingly narrow region of chemical space, she identifies compounds at the periphery that combine promising activity and safety scores with low synthesizability. She manually edits these structures to simplify their synthesis and removes near-duplicate compounds in regions of high structural similarity. The result is a shortlist of diverse, high-scoring antioxidant candidates that reflects both algorithmic optimization and expert judgment for later synthesis and testing, demonstrating the potential of integrating human insight directly into the evolutionary loop.

Another scientist, Bob, focuses on refrigerants for heat pumps; he uses the tool by adapting the default scoring function to his objectives. Not all required properties are available for this use case by default; however, Bob can extend the system via custom API endpoints. Bob adds an endpoint to compute the global warming potential and the cooling capacity [9]. He then defines whether each property should be minimized, maximized, or targeted, and assigns corresponding weights. This scenario showcases how the tool can be flexibly adapted to a wide range of research needs.

[2]Demo available at: https://gems.ivia.ch/.

6 Evaluation

To assess the tool's usability and design, we collected feedback from two domain scientists with minimal experience in generative chemistry models. Each participant completed two tasks: optimizing the default scoring function and steering toward simpler, more biodegradable antioxidant structures, followed by a usability survey (responses are in the supplementary material).

The feedback was overall positive. The tool was considered useful (4.5/5) and well-designed (4/5), but received a passable score for intuitiveness (3.5/5). The configuration interface and constraint features (*e.g.*, exclude substructures) were highly valued. Participants favored the Grid View for inspection and the Table View for filtering and sorting. The Projection View was used less due to cross-view tracking difficulties. This could be solved by a better linking strategy of the views such that users can see their selections and modifications across views. Mutation, crossover, and manual editing were considered valuable for making targeted structural changes. Participants' main difficulty was understanding the genetic algorithm, which required explanation; they suggested adding more onboarding and guidance. Nevertheless, they reported enjoying the interaction and would use GEMS again for their research (4.5/5).

As preliminary evidence for the benefits of expert steering, we compared the top-100 candidates produced by a standard GA run with those of an expert-guided generation. While the standard GA achieved higher scores, the expert-guided candidates were more diverse and markedly more often solvable by computational retrosynthesis (65 vs. 32 of 100). A group of three environmental scientists expressed a slight preference for the expert-evolved candidates in a blind assessment but noted high uncertainties in their assessment due to a lack of experimental data for similar compounds. Details can be found in the Appendix.

7 Conclusion & Future Work

We present GEMS, a human-in-the-loop visual tool for sustainable molecular design that combines algorithmic exploration with direct expert intervention. Unlike in traditional GA workflows, the domain scientist is an active collaborator with direct control over the evolutionary process, enabling them to inspect, edit, and steer evolving populations. Early feedback indicated that participants would use GEMS in their research.

Future work will include adding an onboarding option, conducting an extended user study examining steering strategies, and leveraging implicit feedback from user interactions to train a preference model that reflects expert intuition. Furthermore, displaying scoring explanations could help users calibrate trust in the underlying surrogate models.

Acknowledgments

We thank Stefan Schmid, Cleo Soldini, and Bernadette Mederer for participating in the user study. This work was funded by the Swiss National Science Foundation (SNSF) within the Sinergia project CRSII_222712, as well as the ETH AI Center postdoctoral fellowship awarded to Christina Humer.

References

[1] D. M. Anstine and O. Isayev. Generative Models as an Emerging Paradigm in the Chemical Sciences. *Journal of the American Chemical Society*, 145(16):8736–8750, 2023. doi: 10.1021/jacs.2c13467

[2] B. Bienfait and P. Ertl. JSME: a free molecule editor in JavaScript. *Journal of Cheminformatics*, 5(1):24, 2013. doi: 10.1186/1758-2946-5-24

[3] N. Brown, M. Fiscato, M. H. Segler, and A. C. Vaucher. GuacaMol: Benchmarking Models for de Novo Molecular Design. *Journal of Chemical Information and Modeling*, 59(3):1096–1108, 2019. doi: 10.1021/acs.jcim.8b00839

[4] H. Brunn, G. Arnold, W. Körner, G. Rippen, K. G. Steinhäuser, and I. Valentin. PFAS: forever chemicals—persistent, bioaccumulative and mobile. Reviewing the status and the need for their phase out and remediation of contaminated sites. *Environmental Sciences Europe*, 35(1):20, 2023. doi: 10.1186/s12302-023-00721-8

[5] O.-H. Choung, R. Vianello, M. Segler, N. Stiefl, and J. Jiménez-Luna. Extracting medicinal chemistry intuition via preference machine learning. *Nature Communications*, 14(1):6651, 2023. doi: 10.1038/s41467-023-42242-1

[6] M. Cihan Sorkun, D. Mullaj, J. M. V. A. Koelman, and S. Er. ChemPlot, a Python Library for Chemical Space Visualization. *Chemistry–Methods*, 2(7):e202200005, 2022. doi: 10.1002/cmtd.202200005

[7] K. Fenner and M. Scheringer. The Need for Chemical Simplification As a Logical Consequence of Ever-Increasing Chemical Pollution. *Environmental Science & Technology*, 55(21):14470–14472, 2021. doi: 10.1021/acs.est.1c04903

[8] S. Fujii, Y. Murakami, T. Yoshizawa, S. Ishida, N. Cho, M. Ohta, T. Honma, K. Yoshizoe, M. Sumita, K. Tsuda, and K. Terayama. ChemTSv3: Generalizing Molecular Design via Flexible Search Space Control. 2025. doi: 10.26434/chemrxiv-2025-kdvrt

[9] A. Goldszal, D. Calanzone, V. Taboga, and P.-L. Bacon. Discovery of Sustainable Refrigerants through Physics-Informed RL Fine-Tuning of Sequence Models, 2025. doi: 10.48550/arXiv.2509.19588

[10] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. LineUp: Visual Analysis of Multi-Attribute Rankings. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2277–2286, 2013. doi: 10.1109/TVCG.2013.173

[11] J. He, C. Hua, Y. Wang, and Z. Zheng. Collaborative Intelligence in Sequential Experiments: A Human-in-the-Loop Framework for Drug Discovery. *Information Systems Research*, 2025. doi: 10.1287/isre.2024.1154

[12] C. Humer, H. Heberle, F. Montanari, T. Wolf, F. Huber, R. Henderson, J. Heinrich, and M. Streit. ChemInformatics Model Explorer (CIME): exploratory analysis of chemical model explanations. *Journal of Cheminformatics*, 14(1):21, 2022. doi: 10.1186/s13321-022-00600-z

[13] C. Humer, R. Nicholls, H. Heberle, M. Heckmann, M. Pühringer, T. Wolf, M. Lübbesmeyer, J. Heinrich, J. Hillenbrand, G. Volpin, and M. Streit. CIME4R: Exploring iterative, AI-guided chemical reaction optimization campaigns in their parameter space. *Journal of Cheminformatics*, 16(1):51, 2024. doi: 10.1186/s13321-024-00840-1

[14] M. Jain, T. Deleu, J. Hartford, C.-H. Liu, A. Hernandez-Garcia, and Y. Bengio. GFlowNets for AI-driven scientific discovery. *Digital Discovery*, 2(3):557–577, 2023. doi: 10.1039/D3DD00002H

[15] P. S. Jayasekara, S. K. Skanchy, M. T. Kim, G. Kumaran, B. E. Mugabe, L. E. Woodard, J. Yang, A. J. Zych, and N. L. Kruhlak. Assessing the impact of expert knowledge on ICH M7 (Q)SAR predictions. Is expert review still needed? *Regulatory Toxicology and Pharmacology*, 125:105006, Oct. 2021. doi: 10.1016/j.yrtph.2021.105006

[16] J. H. Jensen. A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space. *Chemical Science*, 10(12):3567–3572, 2019. doi: 10.1039/C8SC05372C

[17] B. Kale, A. Clyde, M. Sun, A. Ramanathan, R. Stevens, and M. E. Papka. ChemoGraph: Interactive Visual Exploration of the Chemical Space. *Computer Graphics Forum*, 42(3):13–24, 2023. doi: 10.1111/cgf.14807

[18] J. Menke, Y. Nahal, E. J. Bjerrum, M. Kabeshov, S. Kaski, and O. Engkvist. Metis: a python-based user interface to collect expert feedback for generative chemistry models. *Journal of Cheminformatics*, 16(1):100, 2024. doi: 10.1186/s13321-024-00892-3

[19] M. Mitchell. *An introduction to genetic algorithms*. Complex adaptive systems. MIT Press, 1996.

[20] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Á. Fernández-Leal. Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review*, 56(4):3005–3054, 2023. doi: 10.1007/s10462-022-10246-w

[21] Y. Nahal, J. Menke, J. Martinelli, M. Heinonen, M. Kabeshov, J. P. Janet, E. Nittinger, O. Engkvist, and S. Kaski. Human-in-the-loop active learning for goal-oriented molecule generation. *Journal of Cheminformatics*, 16(1):138, 2024. doi: 10.1186/s13321-024-00924-y

[22] R. Posthumus, T. Traas, W. Peijnenburg, and E. Hulzebos. External validation of EPIWIN biodegradation models. *SAR and QSAR in Environmental Research*, 16(1-2):135–148, Feb. 2005. doi: 10.1080/10629360412331319899

[23] R. Rodríguez-Pérez and J. Bajorath. Interpretation of Compound Activity Predictions from Complex Machine Learning Models Using Local Approximations and Shapley Values. *Journal of Medicinal Chemistry*, 63(16):8761–8777, 2020. doi: 10.1021/acs.jmedchem.9b01101

[24] D. Rogers and M. Hahn. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010. doi: 10.1021/ci100050t

[25] B. S. Rubin. Bisphenol A: An endocrine disruptor with widespread exposure and multiple effects. *The Journal of Steroid Biochemistry and Molecular Biology*, 127(1-2):27–34, Oct. 2011. doi: 10.1016/j.jsbmb.2011.05.002

[26] C. Seller-Brison, F. Weissbach, K. Jorner, M. Scheringer, and K. Fenner. Hazard Assessment of Antioxidants as Contaminants of Concern. *Environmental Science & Technology Letters*, 2026. doi: 10.1021/acs.estlett.5c01217

[27] I. Sundin, A. Voronov, H. Xiao, K. Papadopoulos, E. J. Bjerrum, M. Heinonen, A. Patronov, S. Kaski, and O. Engkvist. Human-in-the-loop assisted de novo molecular design. *Journal of Cheminformatics*, 14(1):86, 2022. doi: 10.1186/s13321-022-00667-8

[28] A. Tripp and J. M. Hernández-Lobato. Genetic algorithms are strong baselines for molecule generation, 2023. doi: 10.48550/ARXIV.2310.09267

[29] D. C. Vieira, F. S. Paula, L. E. Yaginuma, and G. Fonseca. iMESc – an interactive machine learning app for environmental sciences. *Frontiers in Environmental Science*, 13, 2025. doi: 10.3389/fenvs.2025.1533292

[30] H. Wang, W. Liu, J. Chen, S. Ji, and G. Bi. ParetoGen: Generative Machine Learning Models To Push the Pareto Optimal Frontier of Functionality-Hazard Trade-offs in Per- and Polyfluoroalkyl Substances Green Alternative Designs. *Environmental Science & Technology*, 2026. doi: 10.1021/acs.est.6c00350

[31] H. Wang, M. Skreta, C. T. Ser, W. Gao, L. Kong, F. Strieth-Kalthoff, C. Duan, Y. Zhuang, Y. Yu, Y. Zhu, Y. Du, A. Aspuru-Guzik, K. Neklyudov, and C. Zhang. Efficient Evolutionary Search Over Chemical Space with Large Language Models. In *International Conference on Learning Representations*, 2025.

[32] Y. Wang, J. Dong, Y. Zhou, Y. Cheng, X. Zhao, W. J. G. M. Peijnenburg, M. G. Vijver, K. M. Y. Leung, W. Fan, and F. Wu. Addressing the Data Scarcity Problem in Ecotoxicology via Small Data Machine Learning Methods. *Environmental Science & Technology*, 59(12):5867–5871, 2025. doi: 10.1021/acs.est.5c00510

[33] Y. Yang, Z. Yang, X. Pang, H. Cao, Y. Sun, L. Wang, Z. Zhou, P. Wang, Y. Liang, and Y. Wang. Molecular designing of potential environmentally friendly PFAS based on deep learning and generative models. *Science of The Total Environment*, 953:176095, 2024. doi: 10.1016/j.scitotenv.2024.176095