Explainable Reaction Atlas

Jason Li*
University of Utah

Rodrigo de la Nuez Moraleda†
Nanyang Technological University

Dhruv Meduri‡
University of Utah

Kelin Xia§
Nanyang Technological University

Bei Wang¶
University of Utah

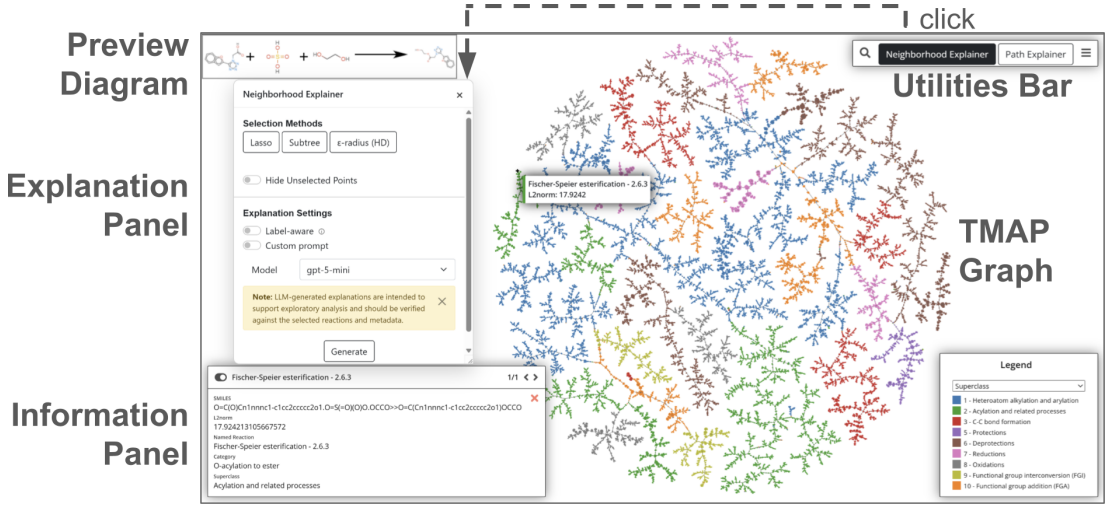


Figure 1: The *Explainable Reaction Atlas* workspace, showing a reaction atlas visualized as an interactive *TMAP Graph* with a force-directed layout. The graph supports zooming, panning, and hover-based inspection of individual reactions. The *Utilities Bar* provides access to (1) *Reaction Search*, (2) the *Neighborhood Explainer*, and (3) the *Path Explainer*; selecting either explainer opens the *Explanation Panel*, which provides LLM-generated explanations to complement human interpretation of the underlying reactions. The legend reflects the current node coloring and can be updated via a drop-down menu to visualize any available categorical or numerical reaction metadata. Selecting a reaction displays its details in the *Information Panel* and its rendered reaction scheme in the *Preview Diagram*.

Abstract

Exploring chemical reaction latent spaces is essential for understanding the organization of complex reaction landscapes. Previous work has shown that transformer-based models can generate high-quality latent representations, known as *reaction fingerprints*, that enable the visualization of reaction spaces as tree-structured atlases. However, interpreting these reaction atlases remains largely manual and time-consuming, limiting their scalability in modern cheminformatics workflows. Although users can readily identify branches, neighborhoods, and transitions, understanding their underlying chemical significance still requires examining numerous individual reactions. We present a framework that combines interactive reaction-atlas visualization with large language model (LLM)-based explanations. Building on reaction atlases, our approach introduces two complementary LLM-driven components: a *Neighborhood Explainer*, which summarizes and compares local regions of the atlas, and a *Path Explainer*, which analyzes transitions of reactions with similar latent space representations along paths defined by user-selected endpoints. The framework supports both label-free and label-aware settings, operating either directly on reaction SMILES (Simplified Molecular Input Line Entry System) representations or by incorporating reaction class labels. We demonstrate the framework on the Schneider 50K dataset, a curated benchmark comprising approximately 50,000 atom-mapped chemical reactions, and show that explanation-guided exploration enables users to interpret local neighborhoods, compare related branches, analyze reaction-space transitions, and identify locally inconsistent regions within chemical reaction latent spaces.

Index Terms: Explainable AI, high-dimensional data, chemical reaction, topological data analysis, dimensionality reduction

1 Introduction

Navigating chemical space is a fundamental task in cheminformatics, underpinning applications ranging from drug discovery to materials design [19, 14]. However, the number of possible compounds and chemical transformations grows combinatorially with the number of atoms and bonds, making direct exploration intractable [34]. Consequently, chemical space is typically analyzed through compact representations of molecules and reactions, ranging from handcrafted descriptors and molecular fingerprints [35] to learned embeddings that capture structural and functional patterns [9, 41]. While effective for prediction, retrieval, and similarity search, these representations reside in high-dimensional spaces that are inherently difficult for humans to interpret. The challenge is particularly acute for chemical reactions, where similarity depends on a combination of reactant and product structures, transformations, reagents, and conditions. As a result, even informative latent representations often require additional semantic context before their organization becomes meaningful to domain experts.

*e-mail: u1244500@utah.edu
†e-mail: rodrigo002@e.ntu.edu.sg
‡e-mail: u1471195@utah.edu
§e-mail: xiakelin@ntu.edu.sg
¶e-mail: beiwang@sci.utah.edu

Dimensionality reduction techniques help address this challenge by projecting high-dimensional representations into lower-dimensional spaces suitable for visual exploration, transforming abstract chemical data into topological structures that experts can readily navigate [27]. Among these methods, TMAP [30] constructs a tree-structured graph that reveals dense neighborhoods, branching patterns, and extended trajectories (see Section 3 for a technical overview). However, while the visualization exposes the topological organization of the latent space, it does not explain the underlying chemistry. Interpreting a branch or neighborhood typically requires manually examining its constituent reactions, which becomes impractical for large datasets. For example, in reaction atlases constructed from the Schneider 50K dataset—a curated benchmark comprising approximately 50,000 atom-mapped chemical reactions—a single local region may contain hundreds of reactions, making exhaustive inspection infeasible [42]. This creates a gap between visual pattern recognition and chemical understanding: the atlas reveals *where* meaningful chemical patterns exist but not *why* they are chemically coherent.

Recent advances in large language models (LLMs) offer an opportunity to bridge this gap. LLMs have demonstrated strong performance across various chemistry tasks, including property prediction, reaction analysis, chemical reasoning, and information extraction [12, 23, 39]. Beyond task-specific predictions, they can generate concise, human-readable descriptions of chemical reactions and reaction sets [10]. This capability naturally complements visual analytics: analysts identify regions of interest, while the LLM provides chemically grounded explanations that can accelerate exploratory analysis by summarizing large neighborhoods, highlighting distinguishing characteristics, and guiding further investigation without requiring users to inspect every individual reaction.

Building on this idea, we introduce *Explainable Reaction Atlas*, an interactive framework for exploring chemical reaction latent spaces by combining TMAP-based reaction atlases with LLM-generated explanations. Rather than treating the LLM as an autonomous analyst, our framework places the user at the center of the exploration process. Users interactively select regions, paths, or boundaries within the atlas, and the LLM generates explanations from the corresponding reaction SMILES (Simplified Molecular Input Line Entry System) and optional metadata; for our use case, this metadata corresponded to the chemical reaction labels. In this way, the model serves as an interpretable semantic layer over the visualization, helping users translate topological patterns into meaningful chemical insights.

To support this workflow, *Explainable Reaction Atlas* augments reaction atlases with two complementary explanation modules that operate on graph-based selections. Inspired by Explainable Mapper [51], the *Neighborhood Explainer* summarizes a selected region or compares two neighborhoods, enabling efficient local characterization and contrastive analysis. The *Path Explainer* analyzes reactions along user-defined paths to describe transitions, revealing how reaction characteristics evolve across the latent space. Together, these components support both exploratory analysis, in which users seek to understand the chemical meaning of a region, and hypothesis-driven investigation, comparing neighborhoods or examine transitions suggested by the atlas. Our contributions are:

- We present a novel interactive framework that combines reaction atlases with LLM-generated semantic explanations, bridging topological visualization and chemical interpretation. The explanation layer is domain-agnostic and can be readily adapted to other graph-based visualization systems, while also helping analysts identify patterns that may not be captured by existing metadata.
- We introduce two interaction-driven explanation modules—the *Neighborhood Explainer* and the *Path Explainer*—that support scalable summarization, comparison, and transition analysis within chemical reaction latent spaces.
- We demonstrate the effectiveness of the framework through a case study on the Schneider 50K dataset, showing how explanation-guided interaction supports scalable exploration and interpretation of chemical reaction space.

Our code is publicly available via GitHub: `https://github.com/tdavislab/explainable-reaction-atlas`.

2 Related Work

We discuss prior work along three axes relevant to *Explainable Reaction Atlas*: representations of chemical reaction space, visual exploration of high-dimensional reaction representations, and topological visual analytics with explanation support [15, 17, 51].

Representation of reaction space. A substantial body of work focuses on encoding chemical reactions as vector representations to support prediction, retrieval, classification and organization [40, 21, 45]. Reaction-space analysis depends heavily on the chosen representation, ranging from handcrafted descriptors and reaction-difference fingerprints to graph-based features and learned neural representations [8, 11, 42]. Learned representations, including transformer-derived reaction fingerprints, have been shown to capture meaningful structure that supports clustering and similarity-based exploration [42]. At the same time, methods such as DRFP show that non-neural, data-independent representations can remain competitive for reaction classification and yield prediction [31].

Visualization of representations. Given these representations, a second line of work focuses on making high-dimensional chemical spaces accessible through visual exploration. Chemical-space maps have long been used to support interactive browsing of large molecular databases, including map-based interfaces for DrugBank, ChEMBL, PubChem, and related compound collections [4, 3]. More recent visual analytics systems combine dimensionality reduction, coordinated molecular views, and interaction mechanisms to help experts inspect molecular similarity, compare representations, and navigate related compounds [37, 16]. In reaction space analysis, low-dimensional projections and atlas-based visualizations have similarly been used to expose clustering, similarity, and global organization of reaction classes [2, 42].

Our work builds directly on the reaction atlases introduced by Schwaller et al. [42], specifically the TMAP-based atlas generated for the Schneider 50K reaction dataset, which serves as the concrete use case and methodological foundation for our framework. In that setting, reaction SMILES are encoded as 256-dimensional transformer-derived reaction fingerprints, and TMAP [30] is used to organize the resulting latent space into an interactive tree-like atlas. The Schneider 50K atlas reveals that reactions from the same class, and often chemically related subclasses, occupy coherent and usually adjacent branches, enabling visual clustering, similarity search, and retrieval of related reactions. However, while this and related visualizations make the structure of reaction space visually accessible, interpreting the chemical meaning of a selected branch, neighborhood, or path still requires manual inspection of many reactions.

Explainable Reaction Atlas extends this specific reaction-atlas setting by adding LLM-assisted neighborhood and path explanations on top of the TMAP structure.

Explainability meets topology. From an explainability perspective, our work builds on visual analytics approaches that use topology as an intermediate representation for interpreting high-dimensional spaces. A prominent example is Mapper [43], which summarizes the shape of high-dimensional data as a graph whose nodes and edges capture regions, branches, and connectivity patterns that are difficult to discern from point-based projections alone. Building on this idea, subsequent systems have transformed topological summaries into interactive analysis tools. Mapper Interactive [53] enables exploratory analysis of high-dimensional datasets through configurable Mapper graphs, while TopoAct [32] and

TopoBERT [33] extend similar topological abstractions to learned representations in neural networks and transformer models. Chemical Mapper [22] uses Mapper graphs to visually explore latent spaces of molecular representations, revealing molecular groupings, relationships, and structural and functional patterns. Collectively, these approaches demonstrate that topology provides an effective abstraction for making latent-space structure inspectable. However, understanding the semantic meaning of individual graph elements—such as branches, neighborhoods, and transitions—still largely depends on manual analysis.

Recent work has begun to bridge this gap by augmenting topological summaries with explanation mechanisms. Mountaineer [44] leverages topology to compare local explanations across machine learning models, while Explainable Mapper [51] introduces LLM-based explanation and verification agents for interpreting Mapper nodes, edges, paths, and trajectories. More broadly, the emerging field of LLM-assisted visual analytics investigates how language models can support explanation, summarization, and human-in-the-loop analysis through systems such as PhenoFlow [17]; comprehensive surveys are provided by [15, 24].

Explainable Reaction Atlas builds on this line of research by bringing LLM-assisted explanation to chemical reaction latent spaces represented as TMAP-based reaction atlases rather than Mapper graphs. Specifically, our framework semantically interprets graph structures—including neighborhoods, branches, and paths—in chemically meaningful concepts derived from reaction SMILES, reaction classes, reagents, functional group transformations, and path-level transitions, enabling users to relate visual patterns to their underlying chemistry.

3 Technical Background on TMAP

TMAP is a graph-based visualization algorithm introduced by Probst et al. [30] for preserving both global and local structure in large, high-dimensional datasets containing up to millions of data points. Rather than directly optimizing a low-dimensional embedding or computing all pairwise distances, TMAP constructs an approximate neighborhood graph and visualizes its underlying tree structure in two dimensions.

Given a high-dimensional point cloud, TMAP first indexes the input vectors using an LSH Forest [5]. For weighted or floating-point vectors, it employs weighted MinHash-style sketches together with consistent weighted sampling to estimate generalized Jaccard similarity from compact hash signatures, enabling efficient approximate nearest-neighbor search [50]. The LSH Forest is parameterized by the number of hash functions (d) used to encode the data and the number of prefix trees (l). Increasing l generally reduces query time at the cost of higher memory consumption.

Using the LSH Forest, TMAP constructs an undirected weighted c-approximate k-nearest-neighbor graph (c-k-NNG), where each vertex represents a data point and weighted edges encode approximate similarity in the original feature space. This step is controlled by the number of nearest neighbors (k) and the LSH query expansion factor (k_c). Rather than embedding the full graph, TMAP computes a minimum spanning tree (MST), producing a sparse connected structure that removes cycles while preferentially preserving short, high-similarity edges. The resulting tree compactly summarizes the data: dense regions typically appear as local branches and subbranches, whereas longer paths connect neighboring regions through the approximate nearest-neighbor structure.

Finally, the MST is embedded into two dimensions using the Open Graph Drawing Framework (OGDF) spring-electrical layout with multilevel multipole force approximation. The layout is controlled by a point-size parameter (p), which indirectly determines the repulsive force between nodes and thus the overall sparsity of the visualization. Unlike projection-based techniques such as t-SNE or UMAP, TMAP visualizes the tree induced by approximate nearest-neighbor relationships rather than directly preserving pairwise distances. Consequently, Euclidean distances in the rendered layout should not be interpreted as exact distances in the original high-dimensional space.

In this work, *Explainable Reaction Atlas* uses TMAP's tree structure to support human-AI collaborative exploration and explanation of the Schneider 50K chemical reaction latent space. We adopt TMAP over Mapper because it preserves fine-grained neighborhood relationships, enabling direct exploration of individual reactions. In contrast, Mapper emphasizes clustered summaries whose topology depends on user-defined choices such as the lens function, interval overlap, and clustering algorithm. TMAP requires fewer parameter choices and is well suited for interactive exploration of reaction latent spaces while retaining access to individual reactions.

4 Design Requirements and Workspace Overview

Current approaches to exploring chemical reaction latent spaces rely on dimensionality reduction techniques combined with manual inspection by domain experts [2, 8, 42]. However, such visualizations often lack semantic interpretability, and the projection of high-dimensional data into two dimensions can introduce distortions, causing reactions that are not genuinely similar to appear artificially close [2, 18, 30].

To address these limitations, we derive a set of design requirements (**R**) that capture the need for scalable and interpretable exploration of reaction latent spaces, and translate them into corresponding design considerations (**C**). Together, these requirements and considerations guide the development of *Explainable Reaction Atlas* workspace, which supports interactive analysis.

These design requirements are primarily informed by prior systems for exploring and interpreting latent spaces, including Embedding Comparator [6], Explainable Mapper [51], and TopoBERT [33]. The requirements proposed by these systems were derived from expert interviews and systematic reviews of the latent-space visualization and analysis literature. Rather than conducting additional expert interviews, we adapt and extend these requirements to address the unique challenges of chemical reaction latent spaces, drawing on needs identified in the chemical reaction visualization literature and capabilities of existing visualization systems.

We therefore design the system around a mixed-initiative, human–AI collaborative workflow in which users retain control over selecting and validating regions of interest, while the LLM provides semantic abstractions of selections that would otherwise be costly to inspect manually [1]. This division of labor is particularly important for scientific analysis, where generated explanations are intended to support expert judgment rather than replace it.

4.1 Design Requirements and Considerations

We present the design requirements (**R**) together with their corresponding design considerations (**C**).

R1. Interactive selection and inspection of reactions. Users require seamless access to detailed information about individual reactions.

C1. Multi-level, on-demand detail retrieval. Support hover-based previews for lightweight inspection and click-based expansion for complete reaction metadata (e.g., structures, conditions, and properties), following a progressive disclosure strategy that balances responsiveness with information density. Because LLM-generated explanations are intended to support exploratory analysis rather than provide authoritative conclusions, users should be able to inspect the reactions underlying each generated summary to validate observed trends and hypotheses. This capability is particularly important because summaries may obscure outliers or minority reaction patterns within a selected region. Direct access to individual

reactions enables users to assess whether an explanation accurately represents the selection or primarily reflects its dominant motifs.

R2. Intuitive selection of regions of interest. Users should be able to explore meaningful subsets of the latent space without requiring prior knowledge of its structure.

C2. Flexible, structure-agnostic selection mechanisms. Provide multiple selection tools that operate directly on the latent representation, including path-, neighborhood-, and region-based interactions. Complement these tools with visual cues, such as density or clustering indicators, to guide exploration of local structures, transitions, and class boundaries. Each selection mode serves a distinct analytical purpose: lasso selection supports visually guided exploration, subtree selection follows the TMAP topology, and the ε-radius selector retrieves neighborhoods in the original latent space for validation.

R3. Configurable LLM-based summarization and comparison. Users need interpretable summaries and comparisons of selected regions to accelerate exploration and guide further investigation of the latent space.

C3. Prompt-driven, region-level semantic abstraction. Integrate an LLM interface that generates summaries and comparative analyses for user-selected regions, while supporting configurable prompts and templates for task-specific queries (e.g., reaction types, functional group transformations, or reaction trends). Prompts are treated as reusable task specifications that guide the model's analytical focus, evidence requirements, and output format [49]. The LLM is intended to complement, rather than replace, traditional rule-based analysis. While rule-based approaches effectively identify predefined reaction patterns and metadata-derived relationships, the LLM can synthesize semantic relationships across user-selected reactions and identify recurring chemical motifs not explicitly captured by the available metadata. This enables analysts to explore both known reaction characteristics and potentially novel attributes while retaining full control over validating generated explanations against the underlying reaction data.

R4. Identification and analysis of ambiguous or inconsistent regions. Users should be able to detect and investigate irregular or ambiguous areas of the latent space.

C4. Local inconsistency detection and diagnostic. Identify regions exhibiting label disagreement or anomalous neighborhood structure using metrics such as local purity or nearest-neighbor disagreement, and provide tools for inspecting these regions and tracing potential causes (e.g., data noise or model limitations).

R5. Scalable exploration of large reaction spaces. The system should support efficient interaction with large reaction datasets.

C5. Hierarchical aggregation and region-level summarization. Support multi-scale exploration through clustering, aggregation, or sampling. Emphasize region-level summaries over exhaustive inspection of reactions, and employ progressive loading and on-demand computation to maintain interactive performance at scale.

4.2 Workspace Overview

Guided by these design requirements, we develop an interactive workspace that combines visualization with LLM-driven explanation capabilities. The system is implemented as a web application that visualizes chemical reaction latent spaces using a TMAP-based tree layout (Figure 1). The *Utilities Bar* provides access to three core components: *Reaction Search*, the *Neighborhood Explainer*, and the *Path Explainer*. Activating either explainer opens the *Explanation Panel*, where users can select regions of interest, configure LLM prompts, and generate explanations.

Reaction inspection. To support fine-grained analysis (**R1**), users can inspect individual reactions represented as nodes in the *TMAP Graph*. Hovering over a node displays a rendered reaction scheme in the *Preview Diagram*, while clicking reveals detailed information, including the reaction SMILES [48] representation and associated metadata, in the *Information Panel*. Multiple reactions can be selected simultaneously, enabling efficient comparison of individual reactions and validation of patterns identified through higher-level summaries.

Reaction search. Accessible through the magnifying glass icon in the *Utilities Bar*, the *Reaction Search* tool enables users to locate reactions by entering a reaction SMILES string or substring. If the query returns 20 or fewer matches, the corresponding nodes are automatically selected and displayed in the *Information Panel*. For larger result sets, users are prompted to explore the matches as a neighborhood within the *Explanation Panel*, supporting both targeted lookup and broader exploratory analysis.

Neighborhood selection. The *Neighborhood Explainer* provides three complementary methods for selecting regions of interest (**R2**). The *Lasso* method enables freeform visual selection. The *Subtree* method requires users to specify a root node and a seed node; the system then performs a breadth-first search from the seed while restricting traversal to the subtree defined by the root. Finally, the *ε-radius* method retrieves all reactions within a user-specified radius of a selected node in the original high-dimensional latent space.

Path selection. The *Path Explainer* supports path-based exploration (**R2**) by allowing users to specify start and end nodes. The system then highlights the unique shortest path connecting the two nodes, exploiting the tree structure of the TMAP graph.

Explainer settings. To satisfy **R3**, the system provides configurable prompts for both summarization and comparison tasks across neighborhood- and path-based selections. A label-aware mode toggle determines whether reaction class labels are included in the LLM input. Label-aware mode is useful when reliable labels are available and class-aware explanations are desired, whereas label-free mode evaluates whether the selected reactions contain sufficient chemical evidence for interpretation without relying on additional information. Users may also enable a custom prompt option to inspect, modify, and extend the default prompt for more specialized analyses.

To maintain interactive performance while respecting LLM context-length constraints (**R5**), the system applies sampling to large selections. Neighborhoods are sampled uniformly at random, while path sampling preserves the sequential ordering and endpoints of the selected path. The sampling strategy is adapted to the analysis task, balancing computational efficiency with representative coverage of the selected reactions.

5 Data and Implementation

We use the Schneider 50K chemical reaction dataset as the primary data source [40]. Reaction fingerprints are generated using the BERT model fine-tuned by Schwaller et al. [42]. The original dataset contains approximately 50,000 atom-mapped reactions spanning 50 reaction classes grouped into 10 superclasses. Prior to analysis, we remove duplicate reactions and standardize their representations, resulting in a dataset of 43,491 unique reactions. Schneider 50K is well suited for our case study because it provides both reaction SMILES and reaction class labels, enabling us to evaluate LLM-generated explanations in both label-aware and label-free settings using the same underlying reaction selections.

5.1 Latent Space Visualization

Each reaction is represented by a 256-dimensional reaction fingerprint generated using the BERT-based model of Schwaller et al. [42]. For a reaction r, we denote its latent representation by $x \in \mathbb{R}^{256}$. These latent vectors serve as the input to the reaction atlas, whose goal is to reveal neighborhood relationships in the high-dimensional reaction space while preserving sufficient global structure for users to identify branches, clusters, and transitions between reaction families.

Following Schwaller et al. [42], we use TMAP [30] to construct a tree-structured representation of the reaction latent space. Because TMAP generates its two-dimensional visualization using a force-directed graph layout, Euclidean distances between nodes in the rendered view do not directly correspond to distances in the original latent space. Consequently, the primary analytical objects in our interface are graph-based topological structures—including neighborhoods, subtrees, and paths—rather than geometric arrangements in the two-dimensional layout.

This tree representation provides the foundation for our explanation tools. The *Neighborhood Explainer* operates on user-defined graph regions, including lasso selections, subtrees, and high-dimensional ε-radius neighborhoods. The *Path Explainer* analyzes the unique tree path connecting two user-selected reactions. Because both neighborhoods and paths are derived from an approximate minimum spanning tree, they should be interpreted as hypotheses about the organization of reaction space rather than definitive chemical relationships. LLM-generated explanations help users evaluate these hypotheses by assessing whether the selected graph structures correspond to coherent reaction motifs, recurring reagents, or meaningful chemical transformations.

5.2 Implementation Overview

The TMAP layout and metadata are precomputed using a standalone Python script, exported with Faerun [29], and loaded into *Explainable Reaction Atlas* for interactive exploration. *Explainable Reaction Atlas* is implemented as a web application with a Flask backend and JavaScript frontend. Explanations are generated using OpenAI's GPT-5 mini, although the model-agnostic backend supports alternative LLMs.

Precomputation separates costly layout construction from interactive analysis. At runtime, the interface retrieves reaction records for the current selection and sends only these records to the explanation backend. Neighborhoods are represented as unordered reaction sets, whereas paths preserve reaction order. In label-aware mode, class labels accompany reaction SMILES; in label-free mode, labels are omitted. The LLM is not fine-tuned on Schneider 50K; responses depend solely on the prompt, selected reactions, optional labels, and analysis task. This design facilitates transfer to new datasets and models without task-specific retraining or annotated explanation data [25, 46].

Prompt templates, adapted from Explainable Mapper [51], support exploratory analysis by summarizing region-level patterns rather than individual reactions. They guide the LLM toward recurring reaction types, reagents, bond changes, and functional-group transformations while avoiding unsupported mechanistic claims. Comparison prompts emphasize contrastive evidence between selections. Responses use structured JSON containing a concise summary and a small set of keywords, enabling consistent presentation and rapid exploration. Users can refine the default prompt or provide custom prompts for deeper analysis. See the supplement for implementation details and complete prompt templates.

6 Case Study

We demonstrate the capabilities of *Explainable Reaction Atlas* through a case study on the Schneider 50K dataset. The selected examples focus on reaction families with well-established importance in medicinal and process chemistry. In particular, Suzuki–Miyaura coupling and amide bond formation are among the most widely used transformations in drug discovery and synthetic chemistry [7, 26, 28, 36, 52]. Moreover, the Schneider 50K dataset is derived from patent reactions and has become a standard benchmark for large-scale reaction classification and similarity analysis. Together, these characteristics make it well suited for demonstrating the framework on transformations that are both chemically meaningful and clearly organized within the reaction atlas.

6.1 Human-AI Collaborative Workflow

Rather than treating the LLM as an autonomous analyst, *Explainable Reaction Atlas* supports a human–AI visual analytics workflow, illustrated in Figure 2, in which users alternate between observation, hypothesis generation, LLM-assisted interpretation, and verification, progressively refining both their prompts and their understanding of the selected region [1, 38].

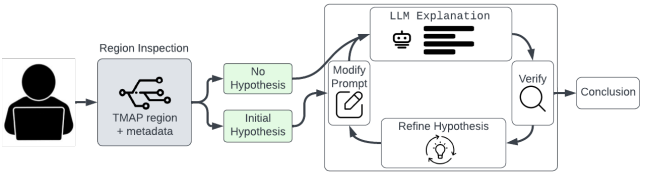


Figure 2: Human–AI collaborative workflow. Users begin by inspecting a region of interest in the TMAP graph and its associated metadata to formulate an initial hypothesis, although exploration may also begin without one. They then generate an LLM-based explanation using either the default or a customized prompt. By comparing the explanation with the underlying reactions, refining their hypothesis, and iteratively adjusting the prompt, users progressively generate and evaluate new explanations until they arrive at a well-supported interpretation of the selected region.

Figure 3 illustrates an example in which a user begins without an initial hypothesis and explores the relationship between two adjacent reaction classes: 8.1.4 (alcohol to aldehyde oxidation, gray) and 1.7.4 (hydroxy to methoxy, blue). The user first applies the default prompt to obtain an overview of the selected region and identify salient chemical features. After verifying these observations against the underlying reactions, the prompt is refined to produce a more targeted explanation, leading to a deeper understanding of the latent-space organization. This cycle of explanation, verification, and prompt refinement can be repeated until the user develops a satisfactory interpretation of the relationship between the two reaction classes. Alternatively, users who begin with an initial hypothesis may bypass the default prompt and directly formulate a customized query.

In this example, the default prompt provides a high-level characterization of the selected neighborhoods, helping the user identify patterns for further investigation. The subsequent analysis examines whether the two reaction classes share chemical similarities beyond the alcohol modifications identified initially. The user refines the prompt to direct the LLM toward additional shared structural and mechanistic features. The resulting explanation identifies several common characteristics, though their prevalence varies substantially: sulfoxides appear in only 20 of 159 reactions and aryl/benzylic motifs in 32, whereas trialkylamine bases occur in 88. This example further illustrates that LLM-generated explanations are intended to guide exploration rather than provide definitive conclusions, underscoring the importance of validating identified patterns against the underlying reaction data.

Figure 4 presents a second example in which the user begins with a specific question about a selected path and therefore customizes the initial prompt rather than relying on the default summary. The goal is to identify transition characteristics beyond the primary distinction between the two reaction classes. In its initial response, the LLM highlights the emergence of sulfonic acid, which manual inspection confirms is present in the final 18 of the 22 reactions and appears shortly before the transition between the two classes. However, verification also reveals that the remaining reported differences are either uncommon or unsupported by the selected reactions: only 2 of the 22 reactions contain halogenated solvents, 5 contain counterions, and none exhibit the functional-group transformations described by the model.

To address these issues, the user refines the prompt to discourage

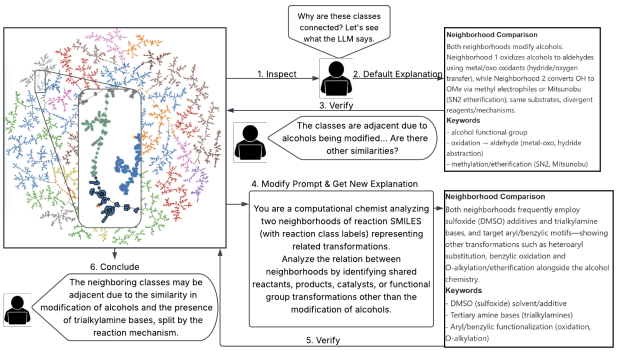


Figure 3: Example neighborhood comparison workflow without an initial hypothesis. Neighborhood 1 corresponds to class 8.1.4 (alcohol-to-aldehyde oxidation, gray), and Neighborhood 2 to class 1.7.4 (hydroxy-to-methoxy, blue).

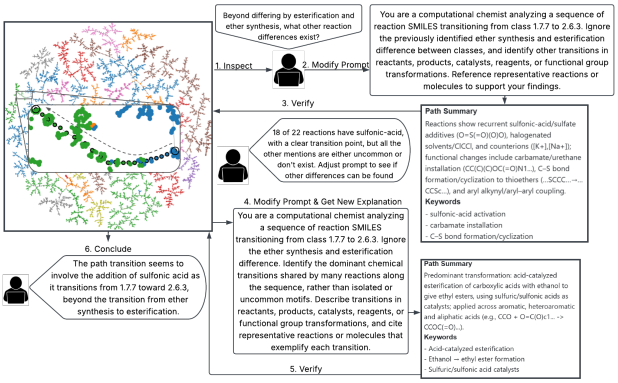


Figure 4: Example of an iterative path summary workflow, beginning with an initial hypothesis. The path begins from class 1.7.7 (Mitsunobu aryl ether synthesis, blue) and transitions to class 2.6.3 (Fischer-Speier esterification, green).

emphasis on isolated motifs and encourage evidence supported by recurring patterns. In the second response, the LLM identifies no additional transition characteristics, instead reaffirming the emergence of sulfonic acid while noting the presence of aromatic and aliphatic acids. Manual verification shows that these latter features occur throughout the path rather than specifically at the transition. The user therefore concludes that, aside from the change in reaction class, the emergence of sulfonic acid is the principal recurring characteristic associated with the selected path transition.

This example illustrates the value of iterative prompt refinement combined with user verification. Rather than treating the initial explanation as authoritative, the analyst uses it to formulate follow-up queries, validate the reported evidence, and refine the interpretation. In this way, the LLM supports exploratory analysis while reducing the risk that uncommon or hallucinated patterns are mistaken for meaningful transition characteristics.

6.2 Neighborhood Explainer

Analysis of local neighborhoods in the reaction atlas reveals that spatially proximate reactions typically belong to the same reaction class. Regions selected with the *Neighborhood Explainer* are therefore usually dominated by a single class, indicating that the learned latent representations organize reaction identity in a geometrically coherent manner. Beyond this local consistency, adjacent branches often correspond to closely related subclasses, suggesting that the latent space also captures higher-order relationships among reaction families. In practice, these neighborhoods enable users to retrieve reactions most similar to a query reaction and explore analogous transformations, while the accompanying LLM-generated summaries help explain the shared chemical characteristics of the retrieved reactions [42].

The value of the *Neighborhood Explainer* extends beyond identifying the dominant reaction class. It also exposes the chemical evidence underlying the grouping, such as recurring electrophiles, nucleophiles, catalysts, or bond-forming patterns. This allows users to move from a purely visual observation—for example, a compact branch or dense neighborhood—to a mechanistic understanding of why those reactions are positioned close together in the latent space.

Figure 5 illustrates this process for two neighboring branches within superclass 3 (C−C bond formation), corresponding to classes 3.1.1 (bromo Suzuki coupling) and 3.1.6 (chloro Suzuki-type coupling). Both neighborhoods consist primarily of Pd-catalyzed Suzuki cross-coupling reactions that form aryl–aryl or C-heteroaryl bonds using organoboron reagents.

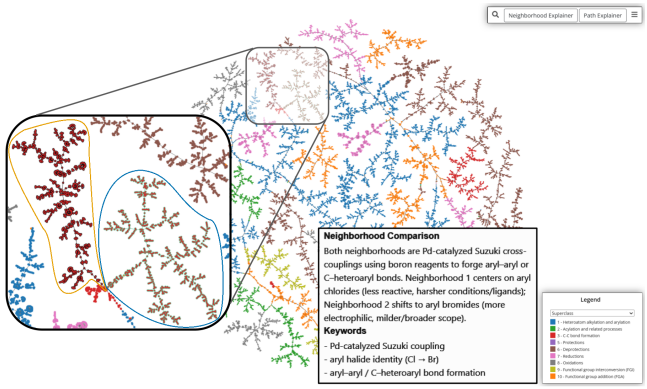


Figure 5: Neighborhood comparison of two Suzuki coupling classes. The zoomed-in view (left) highlights the selected neighborhoods corresponding to class 3.1.1 (bromo Suzuki coupling, circled in orange) and class 3.1.6 (chloro Suzuki-type coupling, circled in blue).

In comparison mode, the *Neighborhood Explainer* identifies the key distinction between the two branches. One is dominated by aryl chlorides, which are less reactive electrophiles and generally require more specialized catalysts and ligands for efficient cross-coupling [20, 26]. The other is dominated by aryl bromides, whose higher reactivity enables milder reaction conditions and a broader substrate scope. Distinguishing reaction families at the level of halide identity would ordinarily require manual inspection of many individual reactions, demonstrating that the explainer supports interpretation at a much finer granularity than broad reaction categories.

This example also illustrates how the system supports iterative hypothesis refinement. From the visualization alone, the two neighboring branches suggest related coupling chemistry. The neighborhood summaries first confirm this hypothesis by identifying both regions as Suzuki-type cross-couplings. The comparative explanation then refines the interpretation by isolating electrophile identity as the principal distinguishing feature: one branch is characterized by aryl bromides, whereas the other is dominated by aryl chlorides. The interaction thus progresses from a visual observation of local proximity to a chemically grounded explanation of how the reaction atlas differentiates closely related cross-coupling classes.

6.3 Path Explainer

The *Path Explainer* supports the analysis of transitions between regions of the reaction atlas. Rather than simply enumerating the reactions along a selected path, it identifies recurring reagents, bond-forming events, functional-group transformations, and mechanistic patterns that characterize the sequence as a whole. This capability is particularly valuable because long paths are difficult to interpret through manual inspection alone: although individual reactions may appear heterogeneous, the ordered sequence can reveal recur-

ring motifs that connect otherwise distinct reaction classes. Consequently, the *Path Explainer* is designed to interpret the chemical relationships encoded by the TMAP topology rather than merely describe the reactions encountered along a path.

It is important to note that paths in the TMAP graph should be interpreted as hypotheses about continuity in the latent space rather than definitive chemical trajectories. Because the graph is derived from an approximate nearest-neighbor structure, some edges may connect reactions that are adjacent in the tree but not truly close in the original high-dimensional space. To help validate proposed transitions, the framework complements path analysis with the ε-radius selector, which retrieves the true high-dimensional neighborhood of reactions along the selected path. The *Path Explainer* then summarizes whether adjacent path segments exhibit coherent reagents, bond changes, or functional-group transformations. A path is considered chemically plausible when both sources of evidence agree: the high-dimensional neighborhoods vary smoothly, and the generated explanation identifies recurring motifs that consistently bridge the selected reaction classes. Conversely, abrupt changes in neighborhood composition or explanations describing unrelated transformations suggest that the observed path may reflect an artifact of the TMAP approximation rather than a meaningful transition through latent space.

Figure 6 illustrates this capability on a path connecting four reaction classes spanning acylation, heteroatom alkylation, functional-group interconversion, and functional-group addition: Ester Schotten–Baumann (2.6.1), Thioether synthesis (1.8.5), Hydroxy-to-chloro (9.1.6), and Chlorination (10.1.2). Despite the chemical diversity of these classes, the *Path Explainer* identifies a coherent sequence of recurring motifs centered on acyl chlorides, sulfuryl- and phosphorus-based chlorinating reagents, and thiol-mediated C–S bond formation. Although thioether synthesis is primarily associated with sulfur chemistry rather than chlorination, the sampled reactions along this segment still involve chlorine-containing reagents, providing a plausible chemical bridge within an otherwise halogenation-dominated transition.

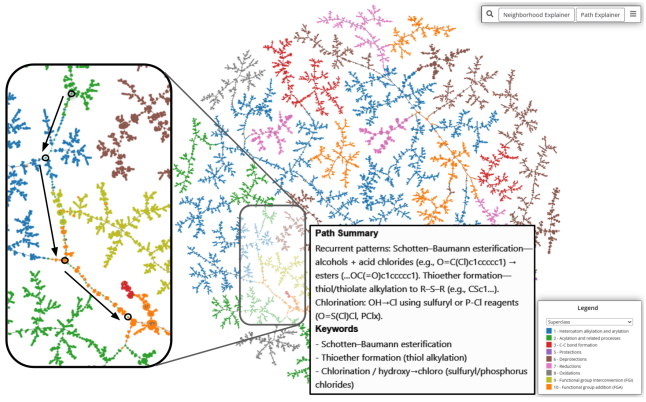


Figure 6: Path explanation across four reaction classes. The zoomed-in view (left) highlights the selected path traversing class 2.6.1 (Ester Schotten–Baumann, green), class 1.8.5 (Thioether synthesis, blue), class 9.1.6 (Hydroxy-to-chloro, olive green), and class 10.1.2 (Chlorination, orange).

This example highlights a use case that differs fundamentally from neighborhood summarization. Rather than asking what characterizes a compact region of the latent space, the user seeks to understand whether an extended connection through the reaction atlas represents a chemically meaningful transition. In this case, the generated explanation suggests that the selected path is not simply an arbitrary sequence of reaction classes, but a progression linked by recurring chlorinated reagents, acylation chemistry, and heteroatom-mediated transformations.

More broadly, this example demonstrates that paths in the reaction atlas can encode interpretable transitions across distinct reaction classes, where shared reactive motifs provide chemically meaningful connections between otherwise disparate regions of the latent space.

6.4 Label-Aware and Label-Free Explanation Modes

A practical consideration for deploying *Explainable Reaction Atlas* is whether reaction class labels are required to produce useful explanations, since many real-world reaction collections are proprietary, newly generated, or only weakly annotated. To investigate this question, both explainers support two complementary operating modes on the same user-selected region. In label-aware mode, reaction class labels are provided alongside the reaction SMILES, whereas in label-free mode the model must infer chemical structure and recurring patterns directly from the SMILES representations. If informative explanations can be generated without class labels, the framework becomes applicable to a broader range of reaction datasets beyond well-curated benchmarks.

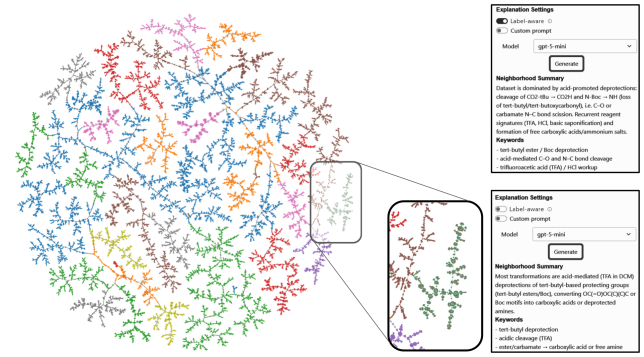


Figure 7: Label-aware and label-free Neighborhood Explainer output on the same selected TMAP region, corresponding to a branch of class 6.2.3 (CO2H-tBu deprotection, brown).

Figure 7 compares the two modes for a branch corresponding to class 6.2.3 (CO_2H-*t*Bu deprotection). Both produce consistent interpretations, identifying acid-mediated removal of tert-butyl protecting groups as the defining transformation while recovering similar reagent signatures, including TFA and HCl. The label-aware mode uses terminology that more closely aligns with the reaction taxonomy, whereas the label-free mode remains grounded in chemical evidence directly inferable from the reaction SMILES. Although class labels improve terminological precision, the comparison demonstrates that meaningful explanations can still be generated without them, supporting the applicability of the framework to unannotated reaction collections.

More broadly, this comparison illustrates the complementary roles of metadata and learned chemical representations. Reaction class labels provide semantic context that stabilizes terminology and facilitates alignment with established taxonomies, whereas the label-free explanations demonstrate that the reaction data themselves often contain sufficient information to support chemically meaningful interpretation.

Figure 8 compares label-aware and label-free explanations for two neighboring reaction classes: class 7.1.1 (Nitro to Amino) and class 10.2.1 (Nitration). Although both modes produce broadly consistent interpretations, the label-aware explanations more closely follow the terminology of the provided reaction class labels, emphasizing the reaction categories rather than the specific reactions within the selected neighborhoods. In contrast, the label-free explanations are grounded more directly in the reaction SMILES, focusing on the recurring reagents, transformations, and chemical patterns present in the selected reactions.

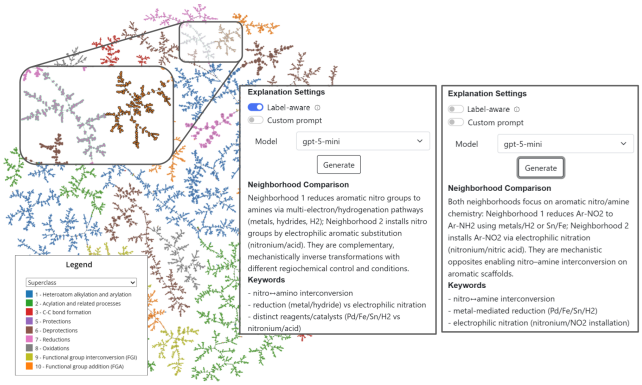


Figure 8: Label-aware and label-free Neighborhood Explainer comparison output on the same selected TMAP regions, corresponding to branches of class 7.1.1 (Nitro to Amino, pink) and class 10.2.1 (Nitration, orange).

This comparison highlights the complementary roles of metadata and reaction representations. Reaction class labels provide semantic context that improves terminological consistency but can also bias the explanation toward predefined taxonomic categories. The label-free mode, by contrast, reveals whether the reaction SMILES alone support the same interpretation. For exploratory analysis, agreement between the two modes increases confidence in the interpretation, whereas disagreement highlights regions where the reaction labels, local chemistry, or generated explanations merit closer inspection.

Figure 9 shows a selected path connecting class 3.1.5 (Bromo Suzuki-type coupling) and class 1.7.9 (Williamson ether synthesis). Unlike the previous example, which uses the *Path Explainer* to summarize a long-range trajectory, this analysis focuses on identifying the local path segment that marks the transition between the two reaction classes.

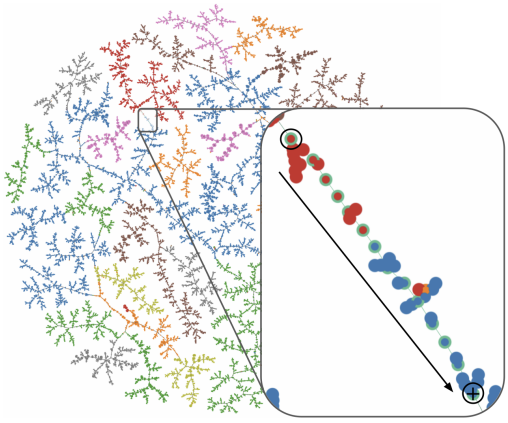


Figure 9: Path selection connecting class 3.1.5 (Bromo Suzuki-type coupling, red) and class 1.7.9 (Williamson ether synthesis, blue).

To support this task, we design a custom prompt that extends the default *Path Explainer* template (see the supplement). Whereas the default prompt summarizes recurring motifs along an ordered path, the custom prompt instructs the model to segment the path into coherent regions, localize the most likely transition point using path indices, and justify the inferred boundary with evidence extracted from the reaction SMILES. This prompt follows established prompt-engineering principles by explicitly specifying the analytical objective, evidence requirements, and desired output structure [49]. By framing the task in this way, the model is encouraged to identify localized transitions rather than produce only a global path summary. In label-aware mode, the prompt further emphasizes that reaction class labels serve as auxiliary context, while the explanation should remain grounded in observable chemical evidence from the selected reactions.

Both the label-aware and label-free outputs identify the selected path as a transition from Suzuki-type cross-coupling chemistry to ether-forming reactions, indicating that the chemical boundary is discernible with or without reaction class labels. In label-aware mode, the explanation describes a shift from boron- and Pd-mediated C–C bond formation to alkoxide-mediated C–O bond formation, using taxonomy-aligned terminology such as Suzuki cross-coupling and Williamson ether synthesis. In label-free mode, the model recovers the same underlying transition directly from the reaction SMILES, identifying the disappearance of $B(OH)_2$/Pd cross-coupling motifs and the emergence of alkoxide and phenoxide nucleophiles associated with C–O bond formation.

We further analyze the path shown in Figure 9 using a custom prompt designed to identify the chemical similarities between the two reaction classes (see the supplement for details), both with and without reaction class labels. In both modes, the model identifies a common theme: reactions involving aryl bromides that form either C–C or C–O bonds. However, the label-free explanation places greater emphasis on coupling with organoboron reagents, a defining feature of class 3.1.5 (Bromo Suzuki-type coupling), which predominates in the first half of the selected path but is absent from class 1.7.9 (Williamson ether synthesis), whose reactions do not involve organoboron reagents.

This example illustrates that, although custom prompts enable targeted exploration, the resulting explanations may still overemphasize localized patterns and therefore benefit from manual verification. In contrast, the label-aware explanation distinguishes between the two reaction classes by identifying aryl bromides as coupling with either organoboron reagents (class 3.1.5) or alkoxide/phenolate nucleophiles (class 1.7.9), producing a more balanced characterization of the transition. While this difference is expected given the additional semantic information available in label-aware mode, it suggests that richer metadata can improve the specificity and completeness of the generated explanations.

6.5 Locally Inconsistent Points

Locally inconsistent points are reactions whose assigned class differs from that of their surrounding neighborhood in the latent space. Identifying these points can help uncover potential misclassifications and guide targeted model refinement [47, 54]. By combining LLM-generated explanations with local neighborhood analysis, users can better understand the chemical characteristics shared by nearby reactions and diagnose the factors contributing to these inconsistencies. In particular, examining the corresponding high-dimensional (HD) neighborhood reveals overlapping chemical features or competing reaction patterns that may explain why a reaction is embedded near a different class.

Unlike the previous examples, this use case focuses on a reaction that is informative precisely because it deviates from its local context. The generated explanation helps distinguish among three possible causes: an incorrect class label, a chemically ambiguous transformation, or a neighborhood artifact introduced by the learned representation. In each case, the explanation serves as a triage mechanism, directing the analyst toward the reactions and chemical features that warrant closer inspection.

Figure 10 demonstrates this workflow using the HD neighborhood selector to retrieve the true latent-space neighborhood of an inconsistent reaction. The LLM-generated explanation identifies a mixture of three reaction classes: 2.1.1 (amide Schotten–Baumann), 10.4.2 (methylation), and 1.6.4 (chloro N-alkylation). Inspection of the retrieved reactions confirms that the explanation correctly attributes the inconsistency to shared reactivity patterns, specifically nucleophilic N/O attack on acid chlorides (R–COCl), often in the presence of bases or salts. It also identifies the defining

chemical characteristics of each reaction class represented within the neighborhood. Together, these results demonstrate that LLM-generated explanations can effectively support the rapid identification and interpretation of local inconsistencies in chemical reaction latent spaces.

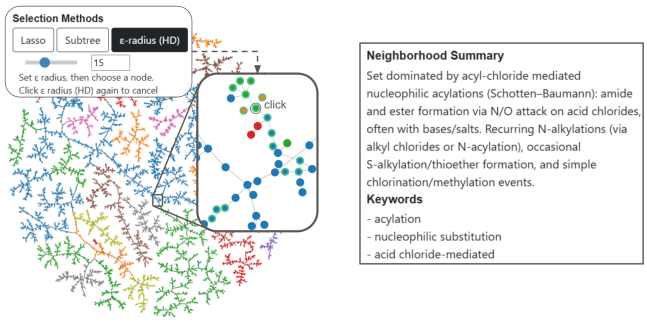


Figure 10: Identification of a locally inconsistent reaction from class 2.1.1 in the latent space. Using the HD Neighborhood selector with an ε-radius of 15 retrieves the reaction's high-dimensional neighborhood, including its TMAP neighbors. The selected neighborhood contains a mixture of three reaction classes: 2.1.1 (amide Schotten–Baumann, green), 10.4.2 (methylation, orange), and 1.6.4 (chloro N-alkylation, blue).

A complementary custom prompt can assess such neighborhoods without assuming that spatial proximity implies chemical coherence. In this example, the model characterizes the neighborhood as only partially consistent: shared acyl chloride reactivity explains much of the grouping, whereas the presence of chemically distinct reactions suggests that the MST may overemphasize this common reactive motif. The full prompt is provided in the supplement.

6.6 Case Study Summary

Overall, the case study demonstrates that the proposed interaction paradigm supports multiple analysis tasks. Neighborhood summaries help users interpret compact local structures, neighborhood comparisons reveal distinctions between adjacent branches, path explanations characterize transitions across the latent space, and inconsistent-point analysis facilitates targeted investigation of ambiguous regions. Across all of these workflows, *Explainable Reaction Atlas* complements rather than replaces manual validation. By transforming large reaction selections into concise, evidence-grounded explanations, it helps users focus their attention on the most relevant patterns while enabling them to verify, challenge, and refine the generated interpretations. See the supplement for a comparison of this human–AI collaborative workflow with current exploration practices.

7 Limitations

Projection distortion. TMAP does not preserve the exact geometry of the reaction latent space. To scale to large datasets, it replaces exact nearest-neighbor search with a c-approximate k-nearest-neighbor graph (c-k-NNG) constructed using an LSH Forest [5] and computes a MST over this approximate graph rather than the complete latent-space graph [30]. A second source of approximation arises from the similarity measure. For floating-point vectors, TMAP employs weighted MinHash (I^2CWS) [50] to estimate generalized Jaccard similarity, which is defined only for non-negative vectors. Because BERT-derived reaction fingerprints contain both positive and negative values, the encoding discards sign information, potentially making representations that differ primarily in their negative components appear more similar than they are. Together, these approximations imply that neighborhoods and paths in the atlas should be interpreted as hypotheses about latent-space organization rather than exact geometric relationships. In our workflow, users validate these hypotheses through high-dimensional neighborhood inspection and LLM-generated explanations, although a systematic characterization of TMAP-induced artifacts remains future work.

LLM explanation reliability. *Explainable Reaction Atlas* is intended to support exploratory analysis rather than replace expert chemical judgment. Inspired by verification agents proposed in Explainable Mapper [51], future versions could incorporate robustness checks based on input-preserving perturbations, such as permuting reactants within reaction SMILES and comparing the resulting explanations. Such methods evaluate explanation stability rather than correctness [13]; consequently, even robust explanations require validation against the underlying reactions. This limitation is particularly evident in label-free mode, where the model must infer chemical structure directly from reaction SMILES. Although our case study shows that useful explanations can be generated without reaction class labels, the model may overemphasize local motifs, produce overly broad summaries, or overlook distinctions that are obvious to domain experts. Accordingly, the generated explanations should be treated as hypotheses that guide inspection rather than authoritative chemical conclusions.

Evaluation scope. Our evaluation consists of a qualitative case study rather than a controlled user study or quantitative benchmark of explanation quality. The presented examples cover the principal interaction modes of the system, including neighborhood summarization, neighborhood comparison, path explanation, label-aware and label-free prompting, and locally inconsistent points. While these examples demonstrate the framework's analytical capabilities, they do not quantify improvements in expert efficiency, trust calibration, or decision quality. Future work should include domain-expert user studies comparing manual inspection with explanation-assisted workflows, together with task-specific metrics that assess whether generated summaries preserve important minority patterns within selected regions.

Scope of the current implementation. The current implementation is tightly coupled to TMAP, which suits reaction atlases by scaling to large high-dimensional datasets while exposing tree-like branches for neighborhood- and path-based exploration. However, this structure constrains the interaction primitives: path explanations rely on unique tree paths, while neighborhood analysis depends on how TMAP relates to the original latent space. Future work should examine whether these explanation mechanisms generalize to alternative representations and topological abstractions. The current implementation also lacks persistent analytical provenance and exportable analysis histories, which could be integrated in future versions.

8 Conclusion

We presented *Explainable Reaction Atlas*, an interactive framework combining TMAP-based reaction atlases with LLM-generated explanations to support exploration and interpretation of chemical reaction latent spaces. Our case study on the Schneider 50K dataset demonstrates that the framework supports neighborhood interpretation, branch comparison, path analysis, label-aware and label-free explanation, discovery of potential reaction attributes absent from existing metadata, and diagnosis of locally inconsistent regions while preserving the analyst's central role in validating underlying reactions. Rather than replacing expert judgment, *Explainable Reaction Atlas* serves as a semantic layer that transforms large reaction selections into concise, inspectable hypotheses, bridging visual pattern recognition and chemical interpretation. Future work will focus on improving explanation provenance and uncertainty, strengthening verification, evaluating the framework with domain experts, and extending it to alternative latent-space representations and datasets.

Acknowledgments

This project was supported in part by the Singapore Ministry of Education (MOE) Tier 2 Academic Research Fund (MOE-T2EP20125-0004) and the U.S. National Science Foundation under Grant IIS-2205418. ChatGPT 5.2 was used solely for grammar correction and language polishing.

References

[1] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz. Guidelines for human-AI interaction. In *Proceedings of the 2019 Computer–Human Interaction Conference on Human Factors in Computing Systems*, pp. 1–13. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300233 3, 5

[2] M. Andronov, M. V. Fedorov, and S. Sosnin. Exploring chemical reaction space with Reaction Difference Fingerprints and Parametric t-SNE. *American Chemistry Society Omega*, 6(45):30743–30751, 2021. doi: 10.1021/acsomega.1c04778 2, 3

[3] M. Awale and J.-L. Reymond. Similarity mapplet: interactive visualization of the directory of useful decoys and chembl in high dimensional chemical spaces. *Journal of Chemical Information and Modeling*, 55(8):1509–1516, 2015. doi: 10.1021/acs.jcim.5b00182 2

[4] M. Awale, R. van Deursen, and J.-L. Reymond. Mqn-mapplet: visualization of chemical space with interactive maps of drugbank, chembl, pubchem, gdb-11, and gdb-13. *Journal of Chemical Information and Modeling*, 53(2):509–518, 2013. doi: 10.1021/ci300513m 2

[5] M. Bawa, T. Condie, and P. Ganesan. LSH forest: self-tuning indexes for similarity search. In *Proceedings of the 14th International conference on World Wide Web*, pp. 651–660. Association for Computing Machinery, New York, 2005. doi: 10.1145/1060745.1060840 3, 9

[6] A. Boggust, B. Carter, and A. Satyanarayan. Embedding comparator: Visualizing differences in global structure and local neighborhoods via small multiples. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*, pp. 746–766. Association for Computing Machinery, New York, 2022. doi: 10.1145/3490099.3511122 3

[7] D. G. Brown and J. Bostrom. Analysis of past and present synthetic methodologies on medicinal chemistry: where have all the new reactions gone? miniperspective. *Journal of medicinal chemistry*, 59(10):4443–4458, 2016. doi: 10.1021/acs.jmedchem.5b01409 5

[8] Y. Ding, B. Qiang, Q. Chen, Y. Liu, L. Zhang, and Z. Liu. Exploring chemical reaction space with machine learning models: representation and feature perspective. *Journal of Chemical Information and Modeling*, 64(8):2955–2970, 2024. doi: 10.1021/acs.jcim.4c00004 2, 3

[9] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing Systems*, vol. 28, pp. 2224–2232. Curran Associates, Inc., Red Hook, 2015. 1

[10] C. Edwards, T. Lai, K. Ros, G. Honke, K. Cho, and H. Ji. Translation between molecules and natural language. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 375–413. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022. doi: 10.18653/v1/2022.emnlp-main.26 2

[11] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, pp. 1263–1272. Proceedings of Machine Learning Research, 2017. doi: 10.48550/arXiv.1704.01212 2

[12] T. Guo, K. Guo, B. Nan, Z. Liang, Z. Guo, N. Chawla, O. Wiest, and X. Zhang. What can Large Language Models do in chemistry? A comprehensive benchmark on eight tasks. In *Advances in Neural Information Processing Systems*, vol. 36, pp. 59662–59688. Curran Associates, Inc., Red Hook, 2023. 2

[13] C. Hsieh, C. Yeh, X. Liu, P. Ravikumar, S. Kim, S. Kumar, and C. Hsieh. Evaluations and methods for explanation through robustness analysis. In *International Conference on Learning Representations (ICLR)*, 2021. doi: 10.48550/arXiv.2006.00442 9

[14] B. Huang and O. A. Von Lilienfeld. Ab initio machine learning in chemical compound space. *Chemical reviews*, 121(16):10001–10036, 2021. doi: 10.1021/acs.chemrev.0c01303 1

[15] M. Hutchinson, R. Jianu, A. Slingsby, and P. Madhyastha. LLM-assisted visual analytics: Opportunities and challenges. In *Computer Graphics and Visual Computing*. The Eurographics Association, London, UK, 2024. doi: 10.2312/cgvc.20241237 2, 3

[16] 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 2

[17] J. Kim, S. Lee, H. Jeon, K.-J. Lee, H.-J. Bae, B. Kim, and J. Seo. Phenoflow: a human-llm driven visual analytics system for exploring large and complex stroke datasets. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):470–480, 2024. doi: 10.1109/TVCG.2024.3456215 2, 3

[18] J. A. Lee and M. Verleysen. Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomputing*, 72(7):1431–1443, 2009. doi: 10.1016/j.neucom.2008.12.017 3

[19] C. Lipinski and A. Hopkins. Navigating chemical space for biology and medicine. *Nature*, 432(7019):855–861, 2004. doi: 10.1038/nature03193 1

[20] A. F. Littke and G. C. Fu. Palladium-catalyzed coupling reactions of aryl chlorides. *Angewandte Chemie International Edition*, 41(22):4176–4211, 2002. doi: 10.1002/1521-3773(20021115)41:22<4176::AID-ANIE4176>3.0.CO;2-U 6

[21] E. K. Mallory, A. Acharya, S. E. Rensi, P. J. Turnbaugh, R. A. Bright, and R. B. Altman. Chemical reaction vector embeddings: towards predicting drug metabolism in the human gut microbiome. In *Proceedings of the Pacific Symposium on Biocomputing*, pp. 56–67. World Scientific Publishing, Singapore, 2018. doi: 10.1142/9789813235533_0006 2

[22] D. Meduri, C.-S. Hu, C. Shen, K. Xia, and B. Wang. Mapping chemical space: Topological data analysis of chemical latent space with Mapper. In H.-K. Ahn, M. Hoffmann, and A. Nayyeri, eds., *42nd International Symposium on Computational Geometry (SoCG 2026)*, vol. 367 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 78:1–78:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2026. doi: 10.4230/LIPIcs.SoCG.2026.78 3

[23] A. Mirza, N. Alampara, S. Kunchapu, M. Ríos-García, B. Emoekabu, A. Krishnan, T. Gupta, M. Schilling-Wilhelmi, M. Okereke, A. Aneesh, et al. A framework for evaluating the chemical knowledge and reasoning abilities of large language models against the expertise of chemists. *Nature Chemistry*, 17(7):1027–1034, 2025. doi: 10.1038/s41557-025-01815-x 2

[24] 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 3

[25] G. Niess, H. Razouk, S. Mandic, and R. Kern. Addressing hallucination in causal Q&A: The efficacy of fine-tuning over prompting in LLMs. In *Proceedings of the Joint Workshop of the 9th Financial Technology and Natural Language Processing (FinNLP), the 6th Financial Narrative Processing (FNP), and the 1st Workshop on Large Language Models for Finance and Legal (LLMFinLegal)*, pp. 253–258. Association for Computational Linguistics, Abu Dhabi, UAE, 2025. 5

[26] P. Orecchia, D. S. Petkova, R. Goetz, F. Rominger, A. S. K. Hashmi, and T. Schaub. Pd-catalysed suzuki–miyaura cross-coupling of aryl chlorides at low catalyst loadings in water for the synthesis of industrially important fungicides. *Green Chemistry*, 23(20):8169–8180, 2021. doi: 10.1039/d1gc02602j 5, 6

[27] A. A. Orlov, T. N. Akhmetshin, D. Horvath, G. Marcou, and A. Varnek. From high dimensions to human insight: exploring dimensionality reduction for chemical space visualization. *Molecular Informatics*, 44(1):e202400265, 2025. doi: 10.1002/minf.202400265 2

[28] V. R. Pattabiraman and J. W. Bode. Rethinking amide bond synthesis.

Nature, 480(7378):471–479, 2011. doi: 10.1038/nature10702 5

[29] D. Probst and J.-L. Reymond. FUn: a framework for interactive visualizations of large, high-dimensional datasets on the web. *Bioinformatics*, 34(8):1433–1435, 2018. doi: 10.1093/bioinformatics/btx760 5

[30] D. Probst and J.-L. Reymond. Visualization of very large high-dimensional data sets as minimum spanning trees. *Journal of Cheminformatics*, 12(1):12, 2020. doi: 10.1186/s13321-020-0416-x 2, 3, 5, 9

[31] D. Probst, P. Schwaller, and J.-L. Reymond. Reaction classification and yield prediction using the differential reaction fingerprint DRFP. *Digital Discovery*, 1(2):91–97, 2022. doi: 10.1039/D1DD00006C 2

[32] A. Rathore, N. Chalapathi, S. Palande, and B. Wang. TopoAct: Visually exploring the shape of activations in deep learning. *Computer Graphics Forum*, 40(1):382–397, 2021. doi: 10.1111/cgf.14195 2

[33] A. Rathore, Y. Zhou, V. Srikumar, and B. Wang. TopoBERT: Exploring the topology of fine-tuned word representations. *Information Visualization*, 22(3):186–208, 2023. doi: 10.1177/14738716231168671 3

[34] J.-L. Reymond, L. Ruddigkeit, L. Blum, and R. Van Deursen. The enumeration of chemical space. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(5):717–733, 2012. doi: 10.1002/wcms.1104 1

[35] D. Rogers and M. Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010. doi: 10.1021/ci100050t 1

[36] S. D. Roughley and A. M. Jordan. The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. *Journal of medicinal chemistry*, 54(10):3451–3479, 2011. doi: 10.1021/jm200187y 5

[37] M. V. Sabando, P. Ulbrich, M. Selzer, J. Byška, J. Mičan, I. Ponzoni, A. J. Soto, M. L. Ganuza, and B. Kozlíková. Chemva: Interactive visual analysis of chemical compound similarity in virtual screening. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):891–901, 2021. doi: 10.1109/TVCG.2020.3030438 2

[38] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, 2014. doi: 10.1109/TVCG.2014.2346481 5

[39] M. Schilling-Wilhelmi, M. Ríos-García, S. Shabih, M. V. Gil, S. Miret, C. T. Koch, J. A. Márquez, and K. M. Jablonka. From text to insight: large language models for chemical data extraction. *Chemical Society Reviews*, 54(3):1125–1150, 2025. doi: 10.1039/d4cs00913d 2

[40] N. Schneider, D. M. Lowe, R. A. Sayle, and G. A. Landrum. Development of a novel fingerprint for chemical reactions and its application to large-scale reaction classification and similarity. *Journal of Chemical Information and Modeling*, 55(1):39–53, 2015. doi: 10.1021/ci5006614 2, 4

[41] P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, and A. A. Lee. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *American Chemical Society Central Science*, 5(9):1572–1583, 2019. doi: 10.1021/acscentsci.9b00576 1

[42] P. Schwaller, D. Probst, A. C. Vaucher, V. H. Nair, D. Kreutter, T. Laino, and J.-L. Reymond. Mapping the space of chemical reactions using attention-based neural networks. *Nature Machine Intelligence*, 3(2):144–152, 2021. doi: 10.1038/s42256-020-00284-w 2, 3, 4, 5, 6

[43] G. Singh, F. Mémoli, G. E. Carlsson, et al. Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *Eurographics Symposium on Point-Based Graphics*, pp. 91–100. The Eurographics Association, Goslar, Germany, 2007. 2

[44] P. Solunke, V. Guardieiro, J. Rulff, P. Xenopoulos, G. Y.-Y. Chan, B. Barr, L. G. Nonato, and C. Silva. Mountaineer: Topology-driven visual analytics for comparing local explanations. *IEEE Transactions on Visualization and Computer Graphics*, 30(12):7763–7775, 2024. doi: 10.1109/TVCG.2024.3418653 3

[45] S. Stocker, G. Csányi, K. Reuter, and J. T. Margraf. Machine learning in chemical reaction space. *Nature Communications*, 11(1):5505, 2020. doi: 10.1038/s41467-020-19267-x 2

[46] M. Szep, D. Rueckert, R. von Eisenhart-Rothe, and F. Hinterwimmer. Fine-tuning large language models with limited data: A survey and practical guide. *Transactions of the Association for Computational Linguistics*, 14:341–377, 2026. doi: 10.1162/tacl.a.627 5

[47] Y. Wang, D. Zou, J. Yi, J. Bailey, X. Ma, and Q. Gu. Improving adversarial robustness requires revisiting misclassified examples. In *International Conference on Learning Representations*, 2019. 8

[48] D. Weininger. SMILES, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36, 1988. doi: 10.1021/ci00057a005 4

[49] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, and D. C. Schmidt. A prompt pattern catalog to enhance prompt engineering with ChatGPT. In *Proceedings of the 30th Conference on Pattern Languages of Programs*, number 5, pp. 1–30. The Hillside Group, Monticello, 2023. doi: 10.64346/PLoP2023p05 4, 8

[50] W. Wu, B. Li, L. Chen, C. Zhang, and P. S. Yu. Improved consistent weighted sampling revisited. *IEEE Transactions on Knowledge and Data Engineering*, 31(12):2332–2345, 2019. doi: 10.1109/TKDE.2018.2876250 3, 9

[51] X. Yan, R. Sevastjanova, S. van der Ben, M. El-Assady, and B. Wang. Explainable Mapper: Charting LLM embedding spaces using perturbation-based explanation and verification agents. *IEEE Transactions on Visualization and Computer Graphics*, 2026. To appear. 2, 3, 5, 9

[52] J. Zhang, N. Sugisawa, K. C. Felton, S. Fuse, and A. A. Lapkin. Multi-objective bayesian optimisation using q-noisy expected hypervolume improvement (q nehvi) for the schotten–baumann reaction. *Reaction Chemistry & Engineering*, 9(3):706–712, 2024. doi: 10.1039/d3re00502j 5

[53] Y. Zhou, N. Chalapathi, A. Rathore, Y. Zhao, and B. Wang. Mapper interactive: A scalable, extendable, and interactive toolbox for the visual exploration of high-dimensional data. In *2021 IEEE 14th Pacific Visualization Symposium (PacificVis)*, pp. 101–110. Institute of Electrical and Electronics Engineers, Piscataway, 2021. doi: 10.1109/PacificVis52677.2021.00021 2

[54] Y. Zhou, Y. Zhou, J. Ding, and B. Wang. Visualizing and analyzing the topology of neuron activations in deep adversarial training. In *Proceedings of ICML Workshop on Topological Data Analysis and Machine Learning (TAGML)*. Proceedings of Machine Learning Research, 2023. 8