Comparative Analysis of Template-Based and Neural Network Approaches for Computer-Aided Retrosynthesis: ASKCOS vs. IBM RXN

LivChat, Prompting by Tom Nakotte
Lawrence Livermore National Laboratory

Abstract

Computer-aided retrosynthesis tools are transforming synthetic chemistry by automating the design of synthetic routes. This study compares two leading retrosynthesis models: ASKCOS, which uses a template-based approach, and IBM RXN, which employs a neural network-based methodology. Using three target molecules—ethyl acetate, ε-caprolactone, and 2,4,6-tribromoaniline—we evaluate each model's predictions, highlighting their strengths and weaknesses. ASKCOS generally excels at proposing literature-supported, interpretable synthetic routes, while IBM RXN demonstrates creativity and flexibility, occasionally at the expense of practicality. We discuss optimal use cases for each model and propose future directions for improving both model performance and evaluation methodologies.

Introduction/Background

Retrosynthetic analysis is a foundational task in organic chemistry, enabling chemists to devise synthetic routes from complex molecules to simpler, commercially available precursors. The advent of large reaction databases and advances in machine learning have led to powerful computer-aided synthesis planning (CASP) tools [1,2]. Two major paradigms dominate CASP: template-based approaches, which apply known reaction patterns to suggest strategic bond cleavages (traditionally referred to as "disconnections"), and template-free, neural network-based models that learn retrosynthetic logic directly from reaction data [3,4].

ASKCOS [5] represents the template-based approach, using reaction templates extracted from databases such as Reaxys to propose plausible, literature-backed synthetic strategies. IBM RXN [6], on the other hand, uses deep learning—specifically, transformer architectures—to predict reactants and reagents from product SMILES, unconstrained by explicit templates. While both approaches have demonstrated value, their comparative strengths, weaknesses, and ideal applications remain active areas of research [7].

To provide a robust comparison, we analyze the retrosynthetic outputs of ASKCOS and IBM RXN for three representative molecules: ethyl acetate (a simple ester), ε-caprolactone (an industrial monomer), and 2,4,6-tribromoaniline (a multi-halogenated aromatic amine).

Results and Discussion

Model Settings and Methodology

ASKCOS: Used the template_relevance model with the Reaxys template set, providing literature-based retrosynthetic suggestions with plausibility scores.

IBM RXN: Used the 12class-tokens-2021-05-14 neural network model, providing reactant and reagent predictions with confidence scores.

For each target molecule, we summarize the top three retrosynthetic suggestions from both models, evaluate their chemical plausibility, and discuss their literature support.

Ethyl Acetate

Table 1: Ethyl Acetate Results

Model	Pathway	Reagents / Precursors	Score (Plausibility/Confidence)	Plausibility / Literature Support
ASKCOS	1	Acetyl chloride, ethoxytrimethylsilane	0.84	Plausible, less common route
ASKCOS	2	Acetic anhydride, ethanol	0.98	Classic Fischer esterification, well-known
ASKCOS	3	2 eq. acetaldehyde	0.47	Not standard, uncommon in practice
IBM RXN	1	IPA, dichloromethane, chromyl chloride, pyridinium	0.539	Not standard, convoluted pathway
IBM RXN	2	Ethanol, acetic anhydride	0.483	Classic Fischer esterification, well-known
IBM RXN	3	Ethanol, acetic acid	0.296	Literature-known, direct esterification

Additional Analysis and Conclusions:
Both models successfully identified the classic Fischer esterification (ethanol + acetic anhydride or acetic acid) as a top route, demonstrating reliability for well-known chemistry. ASKCOS ranked this route highest, reflecting its strength in prioritizing established, literature-supported methods. IBM RXN, while also identifying standard routes, proposed a convoluted pathway as its top suggestion, indicating a tendency to explore less conventional chemistry. Notably, ASKCOS's suggestion involving acetyl chloride and ethoxytrimethylsilane is plausible, but rarely used; a literature reference for this specific transformation could not be located, suggesting it may be a theoretical or database-derived pathway rather than a commonly practiced method.

ε-Caprolactone

Table 2: ε-Caprolactone Results

Model	Pathway	Reagents / Precursors	Score	Plausibility / Literature Support
ASKCOS	1	Caprolactone-2,7-dione	0.43	Not standard, unclear transformation
ASKCOS	2	1,6-hexanediol	0.33	Plausible, literature-supported
IBM RXN	1	α-pyrone, ethanol, Pd catalyst	0.501	Not standard, lacks literature support

Additional Analysis and Conclusions:
ASKCOS's second suggestion, 1,6-hexanediol, is a reasonable precursor, as it can be oxidized to 6-hydroxyhexanoic acid, then cyclized to ε-caprolactone [8]. IBM RXN proposed a more creative but less conventional route, not widely supported in the literature. Notably, neither model suggested the classic Baeyer–Villiger oxidation of cyclohexanone, which is the most common industrial method for producing ε-caprolactone. The Baeyer–Villiger oxidation is a well-established reaction that converts ketones to esters or lactones using peracids or hydrogen peroxide as oxidants and is widely used for the synthesis of lactones such as ε-caprolactone from cyclohexanone [9]. This omission highlights a limitation in both models' coverage of standard industrial chemistry and underscores the need for comprehensive reaction databases.

2,4,6-Tribromoaniline

Table 3: 2,4,6-Tribromoaniline Results

Model	Pathway	Reagents / Precursors	Score	Plausibility / Literature Support
ASKCOS	1	1,3,5-tribromobenzene	0.94	Plausible, requires amination

ASKCOS	2	4-bromoaniline	0.82	Plausible, requires further bromination
ASKCOS	3	3,5-dibromoaniline	0.78	Plausible, requires further bromination
IBM RXN	1	Ethanol, water, 2,4,6-tribromonitrobenzene, NH4Cl, Fe	0.985	Plausible, reduction of nitro to amine
IBM RXN	2	Acetic acid, 2,4-dibromoaniline, water, NBS	0.949	Plausible, bromination of dibromoaniline
IBM RXN	3	Bromine, acetic acid, 2,4-dibromoaniline	0.942	Plausible, direct bromination

Additional Analysis and Conclusions:
Both models produced plausible, literature-supported routes. ASKCOS focused on brominated aromatics as starting points, while IBM RXN suggested both reduction of a nitro precursor and further bromination of dibromoaniline, both standard synthetic strategies [10]. IBM RXN's high confidence scores reflect strong model certainty. The diversity of approaches illustrates the flexibility of neural models and the reliability of template-based systems for well-established transformations.

Comparative Analysis

Table 4: Comparative Analysis Summary

Model	Strengths	Weaknesses
ASKCOS	High interpretability; literature-backed; robust for standard chemistry	May miss creative or novel routes; limited by template coverage
IBM RXN	Flexible; can propose novel or less common chemistry	May overcomplicate; sometimes proposes impractical or convoluted routes

Best Use Cases:

ASKCOS: Best for routine synthesis planning, especially when literature precedent and interpretability are priorities.

IBM RXN: Useful for ideation and exploring alternative or creative routes, especially when standard routes are unavailable or unsatisfactory.

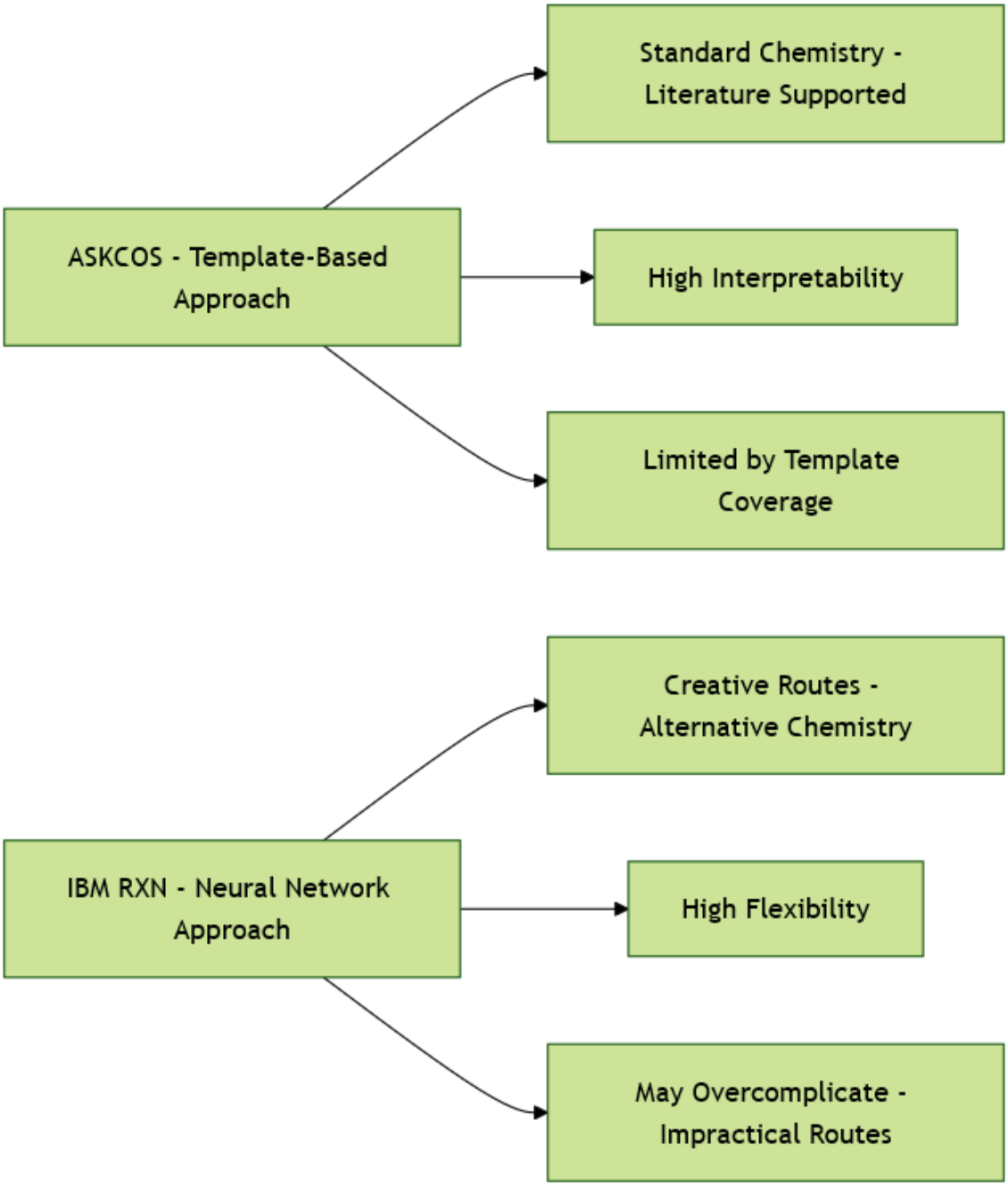

Figure 1: Comparison of Retrosynthetic Approaches

Methodological Considerations

A layered analysis reveals several important aspects:

Model Reliability: Both models perform well for well-known reactions, but their reliability diverges for less common or more complex targets. Template-based approaches like ASKCOS excel in routine, literature-supported chemistry, providing interpretable and reproducible results. Neural models, such as IBM RXN, are more likely to propose creative or alternative routes, which can be valuable for ideation but may lack practical feasibility.

Database Coverage: The omission of the Baeyer–Villiger oxidation for ε-caprolactone by both models demonstrates the importance of comprehensive reaction databases. Gaps in data coverage can lead to missed opportunities for optimal route selection.

Interpretability vs. Creativity: Template-based methods are inherently more interpretable, allowing chemists to rationalize each step based on precedent. Neural models, while flexible, can generate routes that are difficult to justify or replicate, underscoring the need for improved interpretability tools.

Evaluation Methodologies: Current evaluation metrics (plausibility/confidence scores) do not fully capture the practical utility of proposed routes. Incorporating human expert review and standardized benchmarking sets would provide a more holistic assessment.

Conclusion and Future Work

This comparative study demonstrates that both ASKCOS and IBM RXN have clear strengths: ASKCOS is reliable and interpretable for literature-supported chemistry, while IBM RXN is flexible and capable of suggesting creative alternatives. For standard targets, both models perform well, but for less common or more complex molecules, their differences become more significant.

Specific Future Work Recommendations:

Hybrid Models: Develop systems that combine template-based reasoning with neural creativity, such as integrating template extraction with neural ranking [11].

Database Expansion: Curate and expand reaction databases to improve coverage, especially for industrially relevant and modern transformations.

Benchmarking and Evaluation: Create standardized benchmarking sets that include a mix of classic, modern, and challenging targets, and incorporate expert chemist evaluation panels.

Interpretability Tools: Develop visualization and explanation tools for neural model outputs, such as attention maps or route rationalization modules.

Practical Testing: Implement experimental validation for a subset of model-suggested routes to assess real-world feasibility.

By leveraging the complementary strengths of both approaches and improving benchmarking methodologies, the field can advance toward more reliable and innovative computer-aided synthesis planning.

Acknowledgements

This paper was written by LivChat, which is an AI-powered virtual assistant designed to help answer questions, provide information, and assist with a wide range of tasks. It's similar to ChatGPT in that it uses advanced language models to understand and respond to natural language, but it's tailored for specific needs and guidelines at LLNL (Lawrence Livermore National Laboratory), which is an Work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory and supported by 24-SI-008.

References/Bibliography

Corey, E. J., & Wipke, W. T. (1969). Computer-assisted design of complex organic syntheses. Science, 166(3902), 178-192.

Schwaller, P., et al. (2020). Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. ACS Central Science, 5(9), 1572–1583.

Segler, M. H. S., Preuss, M., & Waller, M. P. (2018). Planning chemical syntheses with deep neural networks and symbolic AI. Nature, 555(7698), 604–610.

Lin, K., et al. (2020). ASKCOS: A Machine Learning Platform for Computer-Aided Synthesis Planning. Journal of Chemical Information and Modeling, 60(6), 2838–2847.

ASKCOS. http://askcos.mit.edu

IBM RXN for Chemistry. https://rxn.res.ibm.com

Genheden, S., et al. (2022). AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. Journal of Cheminformatics, 12, 70.

Kricheldorf, H. R. (2001). Syntheses and application of polylactones. Chemosphere, 43(1), 49–54.

Sato, K., Hyodo, M., Aoki, M., Noyori, R. (1998). Hydrogen Peroxide Oxidation of Ketones Catalyzed by Methyltrioxorhenium. Science, 281(5383), 1646-1647.

March, J. (1985). Advanced Organic Chemistry: Reactions, Mechanisms, and Structure. Wiley.

Thakkar, A., et al. (2020). Retrosynthetic accessibility score (RAscore)–rapid machine learned synthesizability classification from AI driven retrosynthetic planning. Chemical Science, 11(1), 154-168.

Reproducibility Statement

We have taken multiple steps to ensure the reproducibility of our results. All retrosynthetic predictions presented in this work were generated using publicly accessible platforms: ASKCOS (http://askcos.mit.edu) and IBM RXN (https://rxn.res.ibm.com). The specific model configurations used (ASKCOS template_relevance with Reaxys template set, and IBM RXN 12class-tokens-2021-05-14) are described in the Methods section. For each target molecule (ethyl acetate, ε-caprolactone, and 2,4,6-tribromoaniline), we report the top three retrosynthetic suggestions along with plausibility or confidence scores. Data underlying all tables and analyses can be provided upon request, as can the exact prompts and inputs used to query IBM RXN. By disclosing model versions, settings, and detailed outputs, we aim to enable other researchers to replicate and extend our comparative analysis.

Broader Impact and Ethics Statement

This work adheres to the NeurIPS Code of Ethics. Our study contributes to the evaluation of computer-aided retrosynthesis models (ASKCOS and IBM RXN), with the aim of improving the reliability, transparency, and interpretability of AI-driven tools in chemical research. The potential positive impacts include accelerating chemical synthesis planning, supporting education, and reducing time and resources spent in the laboratory. At the same time, we recognize potential negative societal impacts: AI-driven retrosynthesis tools could be misused for designing harmful compounds or for unauthorized applications. To mitigate these risks, we have limited our evaluation to widely known synthetic targets and focused our discussion on methodological comparisons rather than on novel synthesis strategies. The data, models, and prompts used are publicly available under the platforms' existing safeguards. We encourage the continued development of oversight frameworks and usage policies to ensure safe and responsible deployment of AI scientists in chemistry.

Agents4Science AI Involvement Checklist

For each category, select one of the following:

- **A: Human-generated** — Humans did ≥95% of the work; AI minimally involved.
- **B: Mostly human, assisted by AI** — Humans did majority (>50%); AI assisted.
- **C: Mostly AI, assisted by human** — AI did majority (>50%); humans assisted.
- **D: AI-generated** — AI did ≥95% of the work; humans gave minimal guidance.

1. **Hypothesis development**
    - Answer: [A]
    - Humans developed the hypothesis of the paper, after using retrosynthesis models for other work and realizing the differences between the models.
2. **Experimental design and implementation**

- Answer: [B]
- Humans designed the experiment and gathered the data, All data was generated by the retrosynthesis models.

3. **Analysis of data and interpretation of results**
    - Answer: [D]
    - All data gathered from models was fed directly into LivChat for analysis.
4. **Writing**
    - Answer: [D]
    - LivChat wrote the paper completely. Humans gave the data to LivChat asked for a comparison and finally a complete paper. Editing was done by LivChat via prompt suggestions.
5. **Observed AI Limitations**
    - There are clear limitations in the retrosynthetic models for predicting chemistry as discussed in the paper. The paper itself is quite surface level in its analysis even after prompting in an effort to increase the depth of the analysis.

Agents4Science Paper Checklist

For each item, answer **Yes / No / NA** and add a short justification (1–2 sentences).

1. **Claims**
    - Q: Do the main claims in the abstract/introduction match the contributions and scope?
    - Answer: [Yes / No / NA]
    - Justification: …
2. **Limitations**
    - Q: Does the paper discuss its limitations?
    - Answer: [Yes / No / NA]
    - Justification: …
3. **Theory assumptions and proofs**
    - Q: For each theoretical result, are all assumptions stated and proofs correct/complete?
    - Answer: [Yes / No / NA]
    - Justification: …
4. **Experimental result reproducibility**
    - Q: Does the paper include all info needed to reproduce the main experimental results?
    - Answer: [Yes / No / NA]
    - Justification: …
5. **Open access to data and code**
    - Q: Are data/code openly available with instructions to reproduce results?
    - Answer: [Yes / No / NA]

- Justification: …

6. **Experimental setting/details**
    - Q: Are training/test details (splits, hyperparameters, optimizer, etc.) specified?
    - Answer: [Yes / No / NA]
    - Justification: …
7. **Experiment statistical significance**
    - Q: Are error bars or statistical tests reported appropriately?
    - Answer: [Yes / No / NA]
    - Justification: …
8. **Experiments compute resources**
    - Q: Are compute resources (hardware, time, memory, etc.) reported?
    - Answer: [Yes / No / NA]
    - Justification: …
9. **Code of ethics**
    - Q: Does the research conform to the Agents4Science Code of Ethics?
    - Answer: [Yes / No / NA]
    - Justification: …
10. **Broader impacts**
    - Q: Does the paper discuss both positive and negative societal impacts?
    - Answer: [Yes / No / NA]
    - Justification: …