

Data-Driven Polyoxometalate Chemistry

Aleksandar Kondinski¹, Nadiia I. Gumerova², Annette Rompel²,
Paolo Falcaro¹, Tobias Schreck³

released: July 6, 2025

¹ Institute of Physical and Theoretical Chemistry
Graz University of Technology
Stremayrgasse 9, 8010 Graz, Austria

² Institut für Biophysikalische Chemie
Fakultät für Chemie, Universität Wien
Josef-Holaubek-Platz 2, 1090 Vienna, Austria

³ Institute of Visual Computing
Graz University of Technology
Inffeldgasse 16c, 8010 Graz, Austria

Preprint No. 1



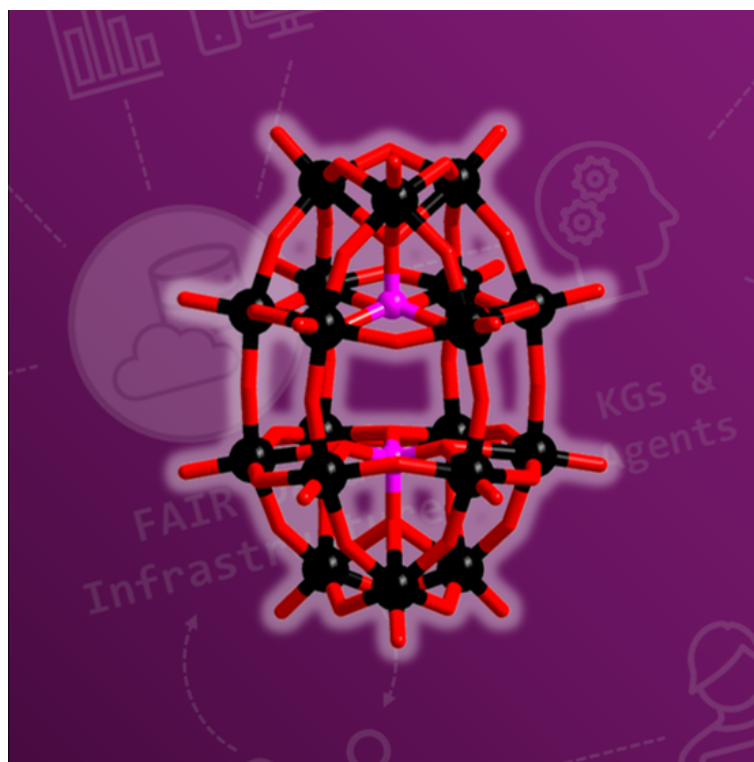
Keywords: Digital Chemistry, Polyoxometalate, Cheminformatics, Artificial Intelligence, FAIR Data

Edited by
Chemical Modelling and Informatics Group
Institute of Physical and Theoretical Chemistry
Graz University of Technology
Stremayrgasse 9, 8010 Graz
Austria (Europe)

E-Mail: kondinski@tugraz.at
World Wide Web: <http://chemin.at/>

Abstract

Polyoxometalates (POMs) are nanoscale, structurally versatile metal–oxo clusters with emerging applications in sustainability, energy, nanoelectronics and life science technologies. Owing to their structural complexity, some all-inorganic POMs are often perceived as serendipitous outcomes from self-assembly processes, which poses challenges for scalable rational design. From this perspective, we therefore examine how the development of POM informatics and, more generally, data-driven POM exploration can pave the way for the molecular engineering of new POM-based materials targeting customised applications. In the process, we highlight recent successes in the digitalisation of POM chemistry and outline which advanced technologies are necessary for progress in this promising area of research.



Highlights

- POM chemistry is entering a digital, data-driven era (POM 4.0).
- Cheminformatics and FAIR data enable machine-actionable POM knowledge.
- AI and robotics accelerate POM discovery and rational design.
- Data-driven frameworks open advances in catalysis, energy, and biomedicine.

Contents

1	Introduction	3
2	Polyoxometalate Cheminformatics	4
3	Current Trends – AI for POMs	6
4	Benefits – AI for “POM-Tech”	7
5	Critical Enablers for POM 4.0	8
6	Outlook	12
	References	13

1 Introduction

Polyoxometalates (POMs) are discrete, highly charged nanomolecular metal-oxo clusters, typically built of early transition metals in high oxidation states.^[42] POMs exhibit a wide range of structural and compositional versatility, ranging from small tri- or tetrametalate clusters to giant assemblies composed of dozens or even hundreds of metal centres.^[43] Their modular composition and tunable oxidation states make them suitable for application in catalysis, molecular electronics, energy storage and life science.^[12,30,39] Moreover, POMs serve as integral building blocks for extended solids and hybrid materials, where they retain their redox-active and structural functionalities.^[62]

POM chemistry has been gradually evolving for over two centuries (see Figure 1). Pioneering chemists such as Scheele and Berzelius had already noted the formation and synthesis of POMs in the late 18th and early 19th centuries; however, their structuring remained unknown until the early 20th century.^[23,49] The foundations of coordination chemistry by Alfred Werner have rekindled interest in this aspect of POMs, motivating various structural proposals and ultimately the first crystallographic elucidations.^[31,49] In this regard, many common “classical” POM archetypical structures such Keggin, Lindqvist, Wells–Dawson, octamolybdate, paratungstate, and decavanadate, were first reported during the 1930s and 1950s, which marks a second stage in the development of POM chemistry.^[31,49] Fundamental interest in POMs as well as applications in catalysis have been driving the molecular engineering of POMs since the 1950s onward.^[8] In this third period of developments, the number of known POMs has significantly expanded. Also, from the early 2000s, advances in computational chemistry and computer technologies have created opportunities for a more comprehensive understanding of the intrinsic stability of POMs, their electronic and spectroscopic properties in relation to experimental observations.^[40] With the rapid digitisation over the past few years and global access to advanced artificial intelligence (AI) tools, POM chemistry and POM-based technologies are gradually entering a new fourth “digital” or “data-driven” stage, to which, inspired by the parallels of the Fourth Industrial Revolution, we may even refer to as “POM 4.0”.^[57]

In this perspective, we, therefore, examine the immediate challenges for data-driven POM chemistry, starting with the necessity to express POM information in a machine-actionable format. This necessity requires the urgent development of polyoxometalate cheminformatics, that is, the computational and informational technique toolset that will enable the solving of polyoxometalate chemical problems. Polyoxometalate cheminformatics will naturally also encompass standardisations capturing the caveats of POM structure representation, the complexity of POM-based processes and the dynamics of POM interconversions. Although applications of AI have enabled advances in POM chemistry in recent years, wider standardisation and semantic structure description enable laying the foundations for automated, evidence-based reasoning as part of broad knowledge systems.^[34] Formalisation of the POM structural and electronic features is also a step for understanding the self-organisation and reactivity of the building block constituents (e.g. multimetal-oxo fragments) that may be similar across different structures or show some correlated activity behaviour (e.g., catalaphores, toxicophores, etc) Finally, accelerating the frontier of data-driven POM chemistry comes with the adaptation and tailoring of AI technologies, aspects that are addressed in this work.

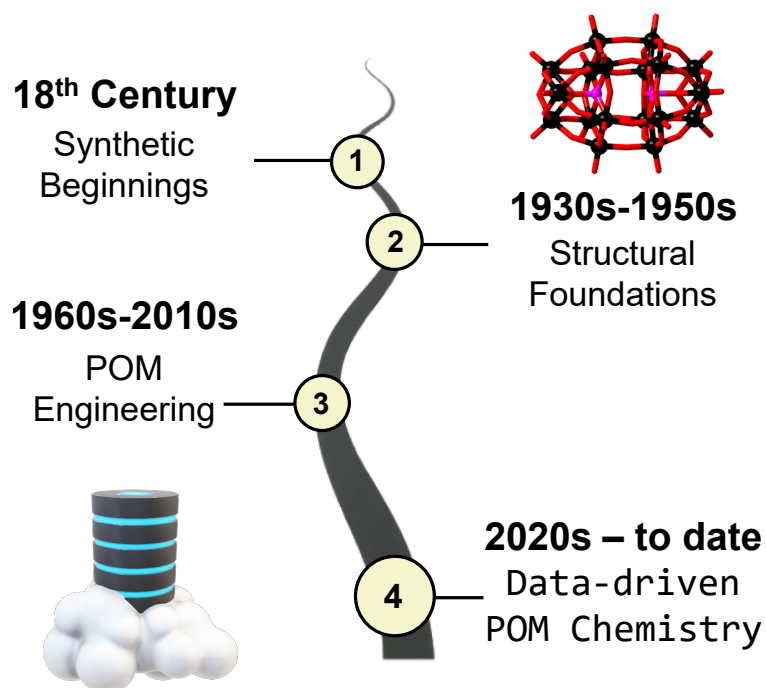


Figure 1: Compact timeline of the evolution of POM chemistry

2 Polyoxometalate Cheminformatics

Information denotes structured data that document the properties of entities, whereas knowledge comprises the logical relations among those data that explain causality.^[34] Robust decision making in chemistry and the symbolic-AI systems that support it therefore require rigorously defined formal representations of chemical concepts and instances, a need that historically drove the emergence of cheminformatics.^[34] POMs as inorganic entities featuring complex coordination and bonding, remain challenging to address *via* common cheminformatics identifiers.^[14] However, much of their subfamily taxonomisation goes along three main factors (see Figure 2.a): i) elemental composition, ii) electronic structure and iii) generic archetype topology. Systematic explorations of POM reactivity require fine tuning of a single factor across a series of POM instances. However, this often may be challenging as the three factors are not fully independent from one another and often influence each other. POMs can be classified according to their elemental composition, specifically by the dominant addenda element (e.g., polyoxotungstates, polyoxomolybdates, polyoxovanadates, etc.). The compositional dimension also provides distinction between a “iso-” or “hetero-”, “mixed-metal” (i.e. “mixed-addenda”), “oxo-substituted”, protonated, “hybrid”, or other type of POM instances.^[6,47] In addition to the elemental aspect, the structural POM archetype refers to a canonical arrangement of the metal–oxo building blocks, which can be seen as analogous to the concept of “frameworks” in the context of zeolite and reticular chemistry.^[48] Well-known examples include the Keggin $[XM_{12}O_{40}]^{n-}$ and Wells–Dawson $[X_2M_{18}O_{62}]^{n-}$ archetypes. Some archetypes

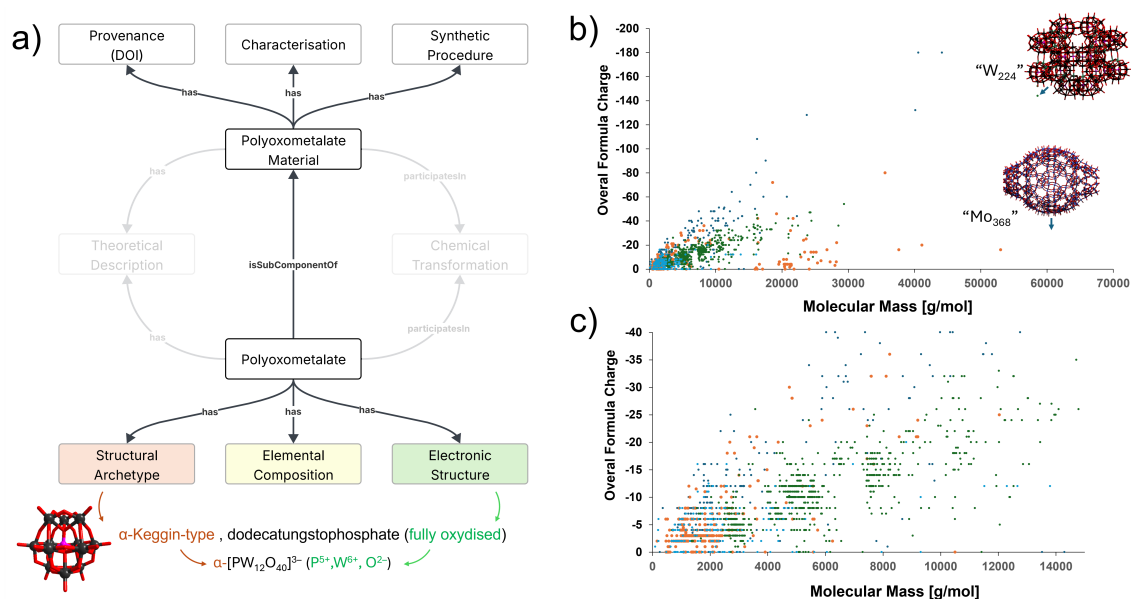


Figure 2: (a) Schematic representation of the herein proposed three key concepts guiding further taxonification of POMs. Note that POMs are entities or subcomponent of typically crystalline POM Materials. Both POMs and POM materials can be a subject of theoretical investigations or considered as components of chemical transformations. (b) A scatter plot from a curated dataset^[35] of POM instance, comparing overall formula charge and molecular mass and a “zoomed in” version with molecular masses up to 30,000 Da. Points are colour-coded by the main addenda metal (e.g., W = green, Mo = orange, V = light blue, and other metals shown in dark blue). Examples of some representative high-nuclearity structures are shown in (b)

can be encountered for different addenda element-based POM, which often allows the construction of mixed metal systems with tailoring properties.^[31] On the other hand, there are also archetypes that appear idiosyncratic and often strictly linked to a particular addenda element or a molecular composition. In more structurally complex POM, different building block combinations may favour the formation of distinct archetypes, depending on reactivity and availability.^[33] The local coordination, which also stipulates the number of terminal oxo/hydroxo/aqua ligands *per* addenda centre, is a reflection of the particular archetype, and it directly affects the structural integrity of the POM to be involved in (multielectron) redox processes. Finally, the electronic structure of the POM, which includes the total overall charge and its distribution, is another important aspect as it provides differentiation between “fully oxidised”, “partially reduced” and fully, highly and super-reduced POMs.^[24] The charge configuration and its (de)localisation can further affect the magnetic and internal bonding properties, as metal–metal bonding can, for some (partially) reduced structures.^[30]

As mentioned earlier, standard chemical identifiers for organic molecules, such as the Simplified Molecular-Input Line-Entry System (SMILES) or the International Chemi-

cal Identifier (InChI), often struggle with complex inorganic species such as POMs.^[14] Features like bridging oxo ligands, protonation equilibria, mixed metal oxidation states, and open-shell electrons are generally difficult to encode in standard string-based representations. The placement of protons on bridging versus terminal oxygens can alter the observed structure considerably, yet conventional encodings lack a robust syntax for capturing these subtle differences. Similarly, partially reduced POMs may exhibit electron delocalisation, variable metal–oxo bond orders, or open-shell spin states. One strategy to address these challenges is to develop extended encodings specifically designed for inorganic systems or to implement hybrid approaches that integrate both geometry and electronic structure. In this regard, the classical *Chemical Markup Language* (CML) which is based on an eXtensible Markup Language (XML) format that can include specialized tags for redox states, bridging ligands, and other elaborate features, may provide a suitable and rather flexible information framework for describing POMs.^[44] However, its verbosity, steeper learning curve, and the need for carefully maintained domain-specific schemas can limit broader adoption.

The recently curated dataset of POM formulas covering the representation of nearly 2000 POM formulas (see Figure 2.b and c), is a step towards the development of POM cheminformatics.^[35] For each POM formula instance, the dataset also incorporates linked information such as literature provenance, total charge, molecular mass, and list of constituting elements, allowing programmatic search and comparisons. Plots of charge versus molecular mass can be indicative of stability limits in super- and undercharged POMs. Future augmentation of this curated POM dataset with electronic information, such as oxidation states of different elements, synthetic information, and reactivity, can be of high interest not only for the development of a comprehensive POM knowledge base, but also for predictive analytics. Among various considerations, the identification and formal representation of POM archetypes is crucial not only for clustering POM instances,^[32] but also for applying rational approaches to understand and discover new ones.

3 Current Trends – AI for POMs

Despite the need for comprehensive POM cheminformatics, AI techniques have been more frequently leveraged at different stages of POM discovery over the past few years. One prominent example is the use of artificial neural networks (ANNs) for modelling POM solubility. Rahman *et al.*^[50] developed an ANN incorporating a physically motivated σ -profile descriptor to capture how alkali cations modulate aqueous solubility across diverse POM salts, revealing three main trends—“normal,” “anomalous,” and an experimentally validated “amphoteric” regime exemplified by $[\text{Ti}_6\text{Nb}_{14}\text{O}_{54}]^{14-}$. Complementing such predictive frameworks, Skjærvø *et al.*^[58] introduced an explainable machine learning (ML) approach, *ML-MotEx*, to interpret time-resolved crystallisation pathways from Pair Distribution Function (PDF) data of transition-metal tungstates, systematically fitting thousands of candidate POM-like fragments and applying SHapley Additive exPlanation (SHAP) to identify the most critical Keggin-type building blocks. Their analysis linked more-ordered POM precursors (e.g. Fe–W compositions) to rapid crystallisation, whereas disordered Co/Ni analogs stalled as amorphous intermediates, thereby highlight-

ing mechanistic connections between local POM motifs and global kinetics. Meanwhile, Anker *et al.*^[5] introduced *POMFinder*, a tree-based eXtreme Gradient Boosting (XGBoost) classifier trained on an extensive simulated POM database, demonstrating 94% accuracy in classifying experimentally measured PDFs, including fast-acquisition or *in situ* data, with SHAP analysis pinpointing metal–metal distances as dominant predictive features. Beyond solubility and crystallisation, Dan *et al.*^[19] showed that unsupervised pipelines can extract hierarchical motifs directly from high-resolution electron microscopy images using a “classify-then-compose” scheme based on rotationally invariant Zernike descriptors. Their algorithm uncovered a novel pentagonal arrangement of metal–oxide octahedra in a Mo–V–Te–Nb mixed-metal oxide, showcasing how motif-level insights help identify hidden structural phases and possible self-assembly routes.

On the other hand, modular robotic platforms have obtained increasing interest in the context of automated discovery and synthesis of POM materials by integrating reconfigurable hardware with sophisticated algorithmic control.^[26,56] These systems employ standardised interfaces and programmable reaction modules that execute precise fluid handling, real-time monitoring, and adaptive condition adjustments based on feedback loops.^[56] Algorithm-driven robots enable high-throughput experimentation across complex synthesis spaces. For instance, an autonomous platform used probabilistic search methods to explore electrochemical routes for inorganic species (including POM clusters) with minimal human intervention.^[37] Reinforcement learning and active learning strategies refine reaction parameters in closed-loop workflows, as exemplified by the discovery of novel POM–MOF structures through iterative screening guided by uncertainty metrics.^[26] Human-in-the-loop approaches can further enhance efficiency. While fully autonomous systems accelerate the identification of “gigantic” POM clusters, collaborative algorithms that combine human intuition with machine learning have outperformed strategies relying solely on either robots or humans.^[20,21] By using adaptive AI models, such as Bayesian optimisation and neural networks, robots can map vast experimental conditions swiftly, refining hypotheses on-the-fly to arrive at optimal or unexpected targets. This synergy between modular robotic architectures and state-of-the-art machine learning frameworks offers a transformative pathway for accelerating POM discovery, ensuring more reproducible, data-rich experimentation and enabling breakthroughs in inorganic materials research that would be exceedingly difficult through manual methods alone.

4 Benefits – AI for “POM-Tech”

The nanoscopic and redox-responsive properties of POMs, make them promising candidates across multiple technological frontiers.^[20,26,63] In catalytic applications, POMs excel at facilitating (photo)oxidations and multi-electron processes, partly due to their strong Brønsted acidity and redox tunability.^[63] However, the coexistence of multiple reactive sites and oxidation states complicates mechanistic elucidation, prompting data-driven approaches to correlate structural motifs with catalytic performance.^[56,67] Neural-network models can predict key descriptors (e.g., deprotonation enthalpies, electron affinity) and thereby pinpoint “catalophores” that promote higher turnover frequencies or improved sta-

bility under reaction conditions.^[46] Recent breakthroughs in interpretable machine learning for oxide-based catalysis have enabled the identification of structural “genes”, providing transparent design rules for active sites.^[67]

Another application area is “POMtronics,” where POM clusters are deployed in molecular electronics and neuromorphic architectures.^[27] Because these clusters can be reversibly oxidized or reduced in molecular junctions, they afford large and stable modulations of conductance for memory and switching functionalities.^[27] Advanced machine-learning techniques (e.g., unsupervised clustering) have proven adept at analysing the extensive current–voltage datasets generated, revealing subtle structure–property correlations that guide device optimisation.^[27] Furthermore, POM-assisted ionic migration can enhance memristor performance, a crucial step toward realising reservoir computing systems.^[66] Drawing on established pipelines from organic electronics^[46], one can explore redox-active motifs and next-generation (hybrid) POM-based electronic devices.

Finally, POMs are emerging as essential components of extended solids and biocomposites, leveraging their robust frameworks and potential bioactivity.^[12,13,18,38,65] Their integration into proteins, polymers, or nanoparticles can yield antibacterial, antiviral, or anticancer effects, as well as enzyme-like activity.^[7,10,11] Notably, POM-protein assemblies have facilitated protein crystallography, while regioselective oxidative cleavage highlights the artificial-metalloenzyme potential of certain clusters.^[1] Yet the rational design of POM-biocomposites remains challenging, as minor alterations in cluster composition and reaction conditions often produce major shifts in biological outcomes.^[65] Moreover, the solution behaviour of POMs, including partial hydrolysis, decomposition products, and the presence of heavy metals such as tungsten and vanadium, requires thorough toxicity assessment.^[25] Data-driven methods such as biomedical knowledge graphs can clarify structure–activity relationships and streamline toxicity predictions, as demonstrated by advances in drug discovery.^[15]

While machine learning has already produced valuable insights, most current models are trained on narrow and sometimes imbalanced data sets and this limits both accuracy and transferability. Studies on porous framework materials show that performance plateaus at modest data volumes and motivate transfer learning and data augmentation.^[54] Hybrid workflows that combine data driven learners with domain knowledge captured in large scale knowledge graphs widen chemical coverage and improve confidence estimates.^[4] Surveys of language models for chemistry recommend attaching explicit uncertainty metrics when data are sparse.^[28] Together with progress in knowledge engineering these developments indicate that blending symbolic reasoning with machine learning can deliver more reliable predictions for POM chemistry in the near future.^[34]

5 Critical Enablers for POM 4.0

Knowledge Graphs and Large Language Models.

The integration of large language models (LLMs) and knowledge graphs (KGs) has significantly advanced reticular chemistry, particularly in the design and discovery of metal–organic frameworks (MOFs). For example, Zheng *et al.* introduced a “GPT-4

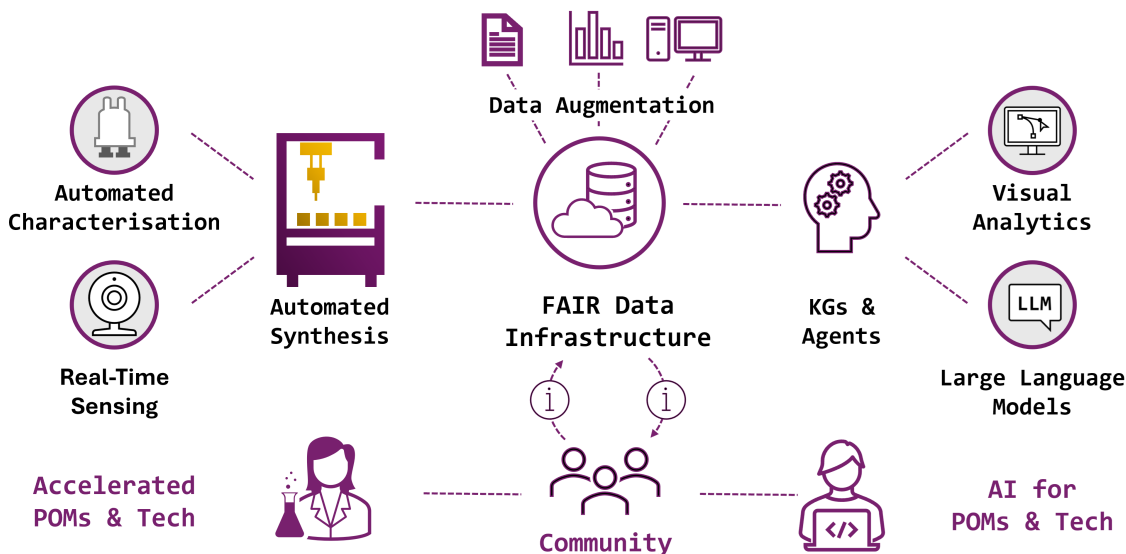


Figure 3: Schematic overview of an envisioned data-driven ecosystem for POM 4.0, with a findable, accessible, interoperable, and reusable (FAIR) data infrastructure at its core. Automated synthesis and real-time characterisation feed experimental data into centralised repositories, enabling data augmentation and knowledge graph (KG)–driven analyses. Large Language Models (LLMs) and AI agents interpret this information, guided by visual analytics tools that engage researchers in the loop.

Reticular Chemist” framework that operates through a cooperative AI–human workflow for planning reactions, transferring solutions, and characterising results, successfully guiding the discovery of an isorecticular series of MOFs.^[68] Likewise, An *et al.* developed the MOF Knowledge Graph (MOF-KG) alongside a benchmark dataset for Knowledge Graph Question Answering in Materials Science (KGQA4MAT), enabling domain experts to query complex chemical data in a natural language format.^[3] These studies highlight the synergy between large language models (LLMs) and knowledge graphs (KGs). LLMs are well-suited for interpreting unstructured text and generating hypotheses, while KGs provide a structured and consistent framework that supports data accessibility and precise querying. In addition, LLMs can assist in formulating complex queries, particularly when users are not fully familiar with the structure of the knowledge graph.

Beyond reticular chemistry, similar integrations have catalysed progress in a broad range of chemical fields. ChemCrow, an LLM-based chemistry agent, demonstrates autonomous performance in organic synthesis, drug discovery, and materials design by coupling natural language generation with specialised toolkits.^[16] Knowledge graph approaches, including those as part of The World Avatar Project, have also proven effective for metal–organic polyhedra and zeolitic frameworks, underscoring their versatility.^[36,52,53] For POM research, domain-specific KGs that capture detailed experimental and computational information can serve as a robust backbone for AI-driven workflows, while LLMs can interpret existing literature, guide new experiments, and help verify re-

sults. This synergy between unstructured data processing and structured knowledge representation holds considerable promise for accelerating POM discovery and innovation.

Computer Vision and Interactive Data Visualisation

Although computer vision may not be regarded as indispensable for advanced POM research, a substantial amount of published POM data is locked in graphical formats. Recently, large-language models with vision capabilities (e.g., GPT-4V) have demonstrated the ability to parse and extract experimental data from figures—including thermogravimetric analysis curves and redox states—in the context of MOFs.^[69] Adapting such vision models to the distinct challenges of POMs, where reaction progress may hinge on subtle colour changes, precipitation onset, or intermediate formation, is a promising yet underexplored research direction.

Many modular platforms for POM synthesis rely on visual cues—such as turbidity or precipitate formation—to guide key steps like filtration. Recent advances from the Hein group demonstrate how webcam-based systems, driven by deep learning, can monitor reactions in real time by detecting colour changes linked to redox states or precipitation events.^[22] Automated platforms (e.g., HeinSight 2.0) can thus promptly identify turbidity or crystallisation endpoints, allowing for immediate intervention. In POM chemistry, where tight control over reaction variables is essential, AI-driven vision can be integrated with other sensors (e.g., pH and redox potential) to adjust reagent feeds or stirring conditions automatically, reducing researcher workload and enhancing reproducibility.

Interactive data visualisations are powerful tools for exploring high-dimensional and time-series data in chemistry. Incorporating human-in-the-loop workflows supports exploration, insight generation, hypothesis testing, and result confirmation. Visual analytics systems,^[55] for instance, combine data visualisation, machine learning (ML), and user interaction to transform raw data into actionable information. Techniques such as visually comparing subspaces in large datasets^[61] or identifying correlations in chemical^[60] and biomolecular data^[17] can reveal patterns that automated algorithms may miss. Visualisation methods also help researchers interpret and guide machine learning models,^[9,41,45] and they are no longer confined to desktop environments—augmented and virtual reality setups,^[51] for example, offer immersive ways to navigate complex molecular assemblies. Collectively, these visual analytics techniques and computer-vision capabilities place humans at the centre of data analysis, bridging the gap between large-scale data pipelines and on-the-fly experimental decision-making in POM research.

Data Infrastructure and Community Efforts

The adoption of FAIR (Findable, Accessible, Interoperable, Reusable) data principles is crucial for advancing POM chemistry.^[64] Currently, many computational results already conform to standardized formats,^[2] which facilitates FAIR data sharing. However, compared to other domains of materials science,^[54] POMs have yet to benefit from large-scale quantum mechanical calculations that generate the extensive datasets needed for data-driven discoveries.

Experimental POM data are typically shared through publications and repository platforms, some of which may not share programmatic access or FAIR data sharing policies, thus requiring intense manual involvement. The growing use of robotic synthetic platforms is likely to change this, as improvements in reproducibility will allow researchers

to report both successful and unsuccessful outcomes, which is essential information for refining machine learning models. Through systematic and transparent reporting of experimental data, POM chemistry can be more effectively integrated into the broader chemical data ecosystem, fostering collaborative innovation and paving the way for data-driven breakthroughs. Journals, funding agencies, and the broader research community all play a crucial role in supporting these efforts.

A practical next step would be a community database that follows the curated and ontology-driven style of ChEBI rather than the broad crowd-sourced model. Every POM entry would carry a persistent identifier plus linked records for synthesis conditions characterisation data electronic state and literature provenance. The schema can extend Chemical Markup Language with terms from the Ontologies for Chemistry initiative^[59] so that records are machine-actionable. Programmatic access through application programming interfaces and Resource Description Framework dumps will feed knowledge graphs and learning algorithms. Open resources such as the Materials Project or the Open Reaction Database have shown that FAIR repositories coupled with agreed ontologies accelerate discovery and power autonomous laboratories.^[2,29] A similar framework would allow POM researchers to capture both successful and null results and turn them into reliable training data for future models.

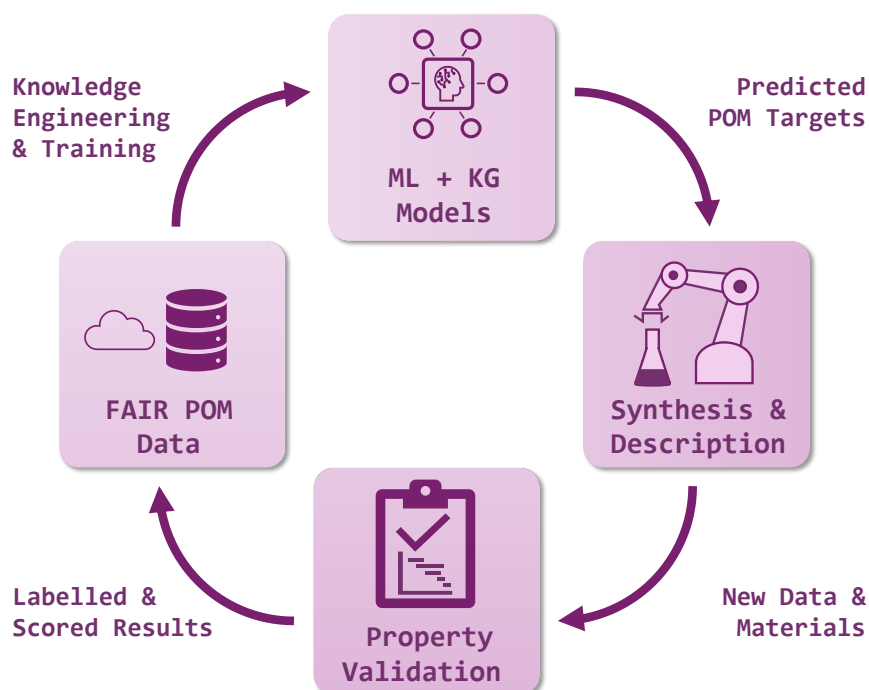


Figure 4: Closed loop workflow for data-driven POM discovery. A FAIR POM database constructs and trains an ML and knowledge graph model which selects targets for robotic synthesis and in situ characterisation. The labelled and uncertainty-scored results return to the database and refine the next cycle.

6 Outlook

Establishing a robust, data-driven ecosystem for POM research will require both refined cheminformatics and flexible, automated experimental workflows. Open-access databases that harmonise experimental and computational insights, complete with consistent descriptors for archetypes, protonation states, and electronic configurations, would enable algorithmic exploration of the POM landscape at an unprecedented scale. In parallel, reconfigurable robotic platforms can integrate *in situ* spectroscopic and scattering methods to capture the subtleties of POM formation, such as slow crystallisation or complex redox processes. By feeding this continuous stream of real-time data into advanced analytics, researchers can systematically map reaction parameters while minimising the trial-and-error approach that often dominates POM synthesis.

Moreover, as summarised in Figure 4, community-driven digital platforms can consolidate these outputs and incorporate AI tools for screening, predicting, and interpreting results, helping to narrow down the vast combinatorial space of potential POM architectures. This approach would also foster a collaborative research culture by encouraging the publication of both successful and partially explored reaction data, an important step toward FAIR-compliant datasets. As computational power increases, tighter integration of quantum chemical simulations within these pipelines will further refine predictions, allowing researchers to anticipate stable structural motifs or redox states before any reagents are mixed. Ultimately, by combining comprehensive data frameworks with responsive automation and targeted simulation, the field can move beyond serendipity and toward genuine, predictive POM design, opening new frontiers in catalysis, energy conversion, and (bio-)materials innovation.

Acknowledgments

AK thanks the Field of Expertise - Advanced Materials Science of TU Graz for Competitive Start-up Funding support (AF22-635-1). This research was funded in part by the Austrian Science Fund (FWF) [DOI 10.55776/P33089 (to A.R.); DOI 10.55776/P33927 (to N.I.G.)]. For open access purposes, the authors have applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

References

- [1] S. A. M. Abdelhameed, F. de Azambuja, T. Vasović, N. D. Savić, T. Ćirković Veličković, and T. N. Parac-Vogt. Regioselective protein oxidative cleavage enabled by enzyme-like recognition of an inorganic metal oxo cluster ligand. *Nat. Commun.*, 14:486, 2023. doi:10.1038/s41467-023-36085-z.
- [2] M. Álvarez-Moreno, C. de Graaf, N. López, F. Maseras, J. M. Poblet, and C. Bo. Managing the computational chemistry big data problem: The iochem-bd platform. *J. Chem. Inf. Model.*, 55(1):95–103, 2015. doi:10.1021/ci500593j. URL <https://doi.org/10.1021/ci500593j>.
- [3] Y. An, J. Greenberg, F. J. Uribe-Romo, D. A. Gómez-Gualdrón, K. Langlois, J. Furst, A. Kalinowski, X. Zhao, and X. Hu. Knowledge graph question answering for materials science (kgqa4mat). In E. Garoufallou and F. Sartori, editors, *Metadata and Semantic Research*, pages 18–29, Cham, 2024. Springer Nature Switzerland.
- [4] M. Anderson, Y. Wu, and D. Fischer. Building a knowledge graph for lithium-ion battery materials: Mining literature at scale. *Nat. Commun.*, 14:1234, 2023. doi:10.1038/s41467-023-02101-6.
- [5] A. S. Anker, E. T. S. Kjær, M. Juelsholt, and K. M. O. Jensen. Pomfinder: identifying polyoxometallate cluster structures from pair distribution function data using explainable machine learning. *J. Appl. Crystallogr.*, 57(1):34–43, 2024. doi:10.1107/S1600576723010014.
- [6] A. V. Anyushin, A. Kondinski, and T. N. Parac-Vogt. Hybrid polyoxometalates as post-functionalization platforms: from fundamentals to emerging applications. *Chem. Soc. Rev.*, 49:382–432, 2020. doi:10.1039/C8CS00854J.
- [7] M. Aureliano, N. I. Gumerova, G. Sciortino, E. Garribba, C. C. McLauchlan, A. Rompel, and D. C. Crans. Polyoxidovanadates’ interactions with proteins: an overview. *Coord. Chem. Rev.*, 454:214344, 2022. doi:10.1016/j.ccr.2021.214344.
- [8] L. C. W. Baker and D. C. Glick. Present general status of understanding of heteropoly electrolytes and a tracing of some major highlights in the history of their elucidation. *Chem. Rev.*, 98(1):3–49, 1998. doi:10.1021/cr960392l.
- [9] E. Beauxis-Aussalet, M. Behrisch, R. Borgo, D. H. Chau, C. Collins, D. S. Ebert, M. El-Assady, A. Endert, D. A. Keim, J. Kohlhammer, D. Oelke, J. Peltonen, M. Riveiro, T. Schreck, H. Strobel, J. J. van Wijk, and T. Rhyne. The role of interactive visualization in fostering trust in AI. *IEEE Comput. Graph. Appl.*, 41(6): 7–12, 2021. doi:10.1109/MCG.2021.3107875.
- [10] A. Bijelic and A. Rompel. The use of polyoxometalates in protein crystallography – an attempt to widen a well-known bottleneck. *Coord. Chem. Rev.*, 299:22–38, 2015. doi:10.1016/j.ccr.2015.03.018.

- [11] A. Bijelic and A. Rompel. Ten good reasons for the use of the tellurium-centered anderson–evans polyoxotungstate in protein crystallography. *Acc. Chem. Res.*, 50(6):1441–1448, 2017. doi:[10.1021/acs.accounts.7b00109](https://doi.org/10.1021/acs.accounts.7b00109).
- [12] A. Bijelic, M. Aureliano, and A. Rompel. The antibacterial activity of polyoxometalates: structures, antibiotic effects and future perspectives. *Chem. Commun.*, 54:1153–1169, 2018. doi:[10.1039/C7CC07549A](https://doi.org/10.1039/C7CC07549A).
- [13] A. Bijelic, M. Aureliano, and A. Rompel. Polyoxometalates as potential next-generation metallodrugs in the combat against cancer. *Angew. Chem. Int. Ed.*, 58(10):2980–2999, 2019. doi:[10.1002/anie.201803868](https://doi.org/10.1002/anie.201803868).
- [14] G. Blanke, J. Brammer, D. Baljovic, N. U. Khan, F. Lange, F. Bänsch, C. A. Tovee, U. Schatzschneider, R. M. Hartshorn, and S. Herres-Pawlis. Making the inchi fair and sustainable while moving to inorganics. *Faraday Discuss.*, 256:503–519, 2025. doi:[10.1039/D4FD00145A](https://doi.org/10.1039/D4FD00145A).
- [15] S. Bonner, I. P. Barrett, C. Ye, R. Swiers, O. Engkvist, A. Bender, C. T. Hoyt, and W. L. Hamilton. A review of biomedical datasets relating to drug discovery: a knowledge graph perspective. *Brief. Bioinform.*, 23(6):bbac404, 2022. doi:[10.1093/bib/bbac404](https://doi.org/10.1093/bib/bbac404).
- [16] A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White, and P. Schwaller. Augmenting large language models with chemistry tools. *Nat. Mach. Intell.*, 6:525–535, 2024. doi:[10.1038/s42256-024-00844-1](https://doi.org/10.1038/s42256-024-00844-1).
- [17] S. Bremm, T. Schreck, P. Boba, S. Held, and K. Hamacher. Computing and visually analyzing mutual information in molecular co-evolution. *BMC Bioinform.*, 11:330, 2010. doi:[10.1186/1471-2105-11-330](https://doi.org/10.1186/1471-2105-11-330). URL <https://doi.org/10.1186/1471-2105-11-330>.
- [18] D. Chang, Y. Li, Y. Chen, X. Wang, D. Zang, and T. Liu. Polyoxometalate-based nanocomposites for antitumor and antibacterial applications. *Nanoscale Adv.*, 4(18):3689–3706, 2022. doi:[10.1039/d2na00391k](https://doi.org/10.1039/d2na00391k).
- [19] J. Dan, X. Zhao, S. Ning, J. Lu, K. P. Loh, Q. He, N. D. Loh, and S. J. Pennycook. Learning Motifs and Their Hierarchies in Atomic Resolution Microscopy. *Sci. Adv.*, 8:eabk1005, 2022. doi:[10.1126/sciadv.abk1005](https://doi.org/10.1126/sciadv.abk1005).
- [20] V. Duros, J. Grizou, W. Xuan, Z. Hosni, D. Long, H. N. Miras, and L. Cronin. Human versus Robots in the Discovery and Crystallization of Gigantic Polyoxometalates. *Angew. Chem. Int. Ed.*, 56:10815–10820, 2017. doi:[10.1002/anie.201705721](https://doi.org/10.1002/anie.201705721).
- [21] V. Duros, J. Grizou, A. Sharma, S. H. M. Mehr, A. Bubliauskas, P. Frei, H. N. Miras, and L. Cronin. Intuition-Enabled Machine Learning Beats the Competition When Joint Human-Robot Teams Perform Inorganic Chemical Experiments. *J. Chem. Inf. Model.*, 59:2664–2671, 2019. doi:[10.1021/acs.jcim.9b00304](https://doi.org/10.1021/acs.jcim.9b00304).

- [22] R. El-khawaldeh, M. Guy, F. Bork, N. Taherimakhssousi, K. N. Jones, J. M. Hawkins, L. Han, R. P. Pritchard, B. A. Cole, S. Monfette, and J. E. Hein. Keeping an “eye” on the experiment: computer vision for real-time monitoring and control. *Chem. Sci.*, 15:1271–1282, 2024. doi:10.1039/D3SC05491H.
- [23] P. Gouzerh and M. Che. From Scheele and Berzelius to Müller: Polyoxometalates (POMs) revisited and the “missing link” between the bottom up and top down approaches. *Actual. Chim.*, 298:9–22, June 2006. URL <https://new.societechimiquedefrance.fr/wp-content/uploads/2019/12/2006-298-juin-Che-p.9.pdf>.
- [24] N. I. Gumerova and A. Rompel. Synthesis, structures and applications of electron-rich polyoxometalates. *Nat. Rev. Chem.*, 2(2):0112, 2018. doi:10.1038/s41570-018-0112.
- [25] N. I. Gumerova and A. Rompel. Speciation atlas of polyoxometalates in aqueous solutions. *Sci. Adv.*, 9(25):eadi0814, 2023. doi:10.1126/sciadv.adi081.
- [26] D. He, Y. Jiang, M. Guillén-Soler, Z. Geary, L. Vizcaíno-Anaya, D. Salley, M. D. C. Gimenez-Lopez, D.-L. Long, and L. Cronin. Algorithm-driven robotic discovery of polyoxometalate-scaffolding metal–organic frameworks. *J. Am. Chem. Soc.*, 146(42):28952–28960, 2024. doi:10.1021/jacs.4c09553.
- [27] C. Huez, D. Guérin, S. Lenfant, F. Volatron, M. Calame, M. L. Perrin, A. Proust, and D. Vuillaume. Redox-controlled conductance of polyoxometalate molecular junctions. *Nanoscale*, 14:13790–13800, 2022. doi:10.1039/D2NR03457C.
- [28] K. M. Jablonka, P. Schwaller, A. Ortega-Guerrero, and et al. Leveraging large language models for predictive chemistry. *Nat. Mach. Intell.*, 6:161–169, 2024. doi:10.1038/s42256-023-00788-1.
- [29] K. Johnson, T. Mori, and M. Garcia. Toward self-driving labs: Automating synthesis with online learning and real-time data acquisition. *ACS Cent. Sci.*, 8:1408–1419, 2022. doi:10.1021/acscentsci.2c00347.
- [30] A. Kondinski. Metal–metal bonds in polyoxometalate chemistry. *Nanoscale*, 13:13574–13592, 2021. doi:10.1039/D1NR02357H.
- [31] A. Kondinski and T. N. Parac-Vogt. Keggin structure, quō vādis? *Front. Chem.*, 6:346, 2018. doi:10.3389/fchem.2018.00346.
- [32] A. Kondinski, A. Menon, D. Nurkowski, M. Farazi, S. Mosbach, J. Akroyd, and M. Kraft. Automated rational design of metal–organic polyhedra. *J. Am. Chem. Soc.*, 144(26):11713–11728, 2022. doi:10.1021/jacs.2c03402.
- [33] A. Kondinski, M. Rasmussen, S. Mangelsen, N. Pienack, V. Simjanoski, C. Nather, D. L. Stares, C. A. Schalley, and W. Bensch. Composition-driven archetype dynamics in polyoxovanadates. *Chem. Sci.*, 13:6397–6412, 2022. doi:10.1039/D2SC01004F.

- [34] A. Kondinski, J. Bai, S. Mosbach, J. Akroyd, and M. Kraft. Knowledge engineering in chemistry: From expert systems to agents of creation. *Acc. Chem. Res.*, 56(2): 128–139, 2023. doi:10.1021/acs.accounts.2c00617.
- [35] A. Kondinski, N. Gumerova, and A. Rompel. A curated polyoxometalate formula dataset and its application to data-driven cluster chemistry. *Data*, 9(11):124, 2024. doi:10.3390/data9110124.
- [36] A. Kondinski, P. Rutkevych, L. Pascazio, D. N. Tran, F. Farazi, S. Ganguly, and M. Kraft. Knowledge graph representation of zeolitic crystalline materials. *Digit. Discov.*, 3:2070–2084, 2024. doi:10.1039/D4DD00166D.
- [37] K. Laws, M. Tze-Kiat Ng, A. Sharma, Y. Jiang, A. J. S. Hammer, and L. Cronin. An Autonomous Electrochemical Discovery Robot that Utilises Probabilistic Algorithms: Probing the Redox Behaviour of Inorganic Materials. *ChemElectroChem*, 11:e202300532, 2024. doi:10.1002/celec.202300532.
- [38] W. Liang, P. Wied, F. Carraro, C. J. Sumby, B. Nidetzky, C.-K. Tsung, P. Falcaro, and C. J. Doonan. Metal–organic framework-based enzyme biocomposites. *Chem. Rev.*, 121(3):1077–1129, 2021. doi:10.1021/acs.chemrev.0c01029.
- [39] R. Liu and C. Streb. Polyoxometalate-single atom catalysts (pom-sacs) in energy research and catalysis. *Adv. Energy Mater.*, 11(25):2101120, 2021. doi:10.1002/aenm.202101120.
- [40] X. López, J. J. Carbó, C. Bo, and J. M. Poblet. Structure, properties and reactivity of polyoxometalates: a theoretical perspective. *Chem. Soc. Rev.*, 41:7537–7571, 2012. doi:10.1039/C2CS35168D.
- [41] L. Meng, S. van den Elzen, and A. Vilanova. Modelwise: Interactive model comparison for model diagnosis, improvement and selection. *Comput. Graph. Forum*, 41(3): 97–108, 2022. doi:10.1111/CGF.14525. URL <https://doi.org/10.1111/cgf.14525>.
- [42] A. Müller and P. Gouzerh. From linking of metal-oxide building blocks in a dynamic library to giant clusters with unique properties and towards adaptive chemistry. *Chem. Soc. Rev.*, 41(22):7431–7463, 2012. doi:10.1039/C2CS35169B.
- [43] A. Müller, P. Kögerler, and A. Dress. Giant metal-oxide-based spheres and their topology: from pentagonal building blocks to keplerates and unusual spin systems. *Coord. Chem. Rev.*, 222(1):193–218, 2001. doi:10.1016/S0010-8545(01)00391-5.
- [44] P. Murray-Rust, H. S. Rzepa, and M. Wright. Development of chemical markup language (cml) as a system for handling complex chemical content. *New J. Chem.*, 25:618–634, 2001. doi:10.1039/B008780G.
- [45] Q. Q. Ngo, F. L. Dennig, D. A. Keim, and M. Sedlmair. Machine learning meets visualization - experiences and lessons learned. *it Inf. Technol.*, 64(4-5):169–180, 2022. doi:10.1515/ITIT-2022-0034.

- [46] D. M. Packwood, Y. Kaneko, D. Ikeda, and M. Ohno. An intelligent, user-inclusive pipeline for organic semiconductor design. *Adv. Theory Simul.*, 6:2300159, 2023. doi:[10.1002/adts.202300159](https://doi.org/10.1002/adts.202300159).
- [47] M. T. Pope. *Heteropoly and Isopoly Oxometalates*, volume 8 of *Inorganic Chemistry Concepts*. Springer Berlin, Heidelberg, 1983. ISBN 978-3-662-12006-4. ISBN 978-3-662-12006-4.
- [48] M. T. Pope and A. Müller. Polyoxometalate chemistry: An old field with new dimensions in several disciplines. *Angew. Chem. Int. Ed. Engl.*, 30:34–48, 1991. doi:[10.1002/anie.199100341](https://doi.org/10.1002/anie.199100341).
- [49] R. Pütt, A. Kondinski, and K. Y. Monakhov. Die „klassischen“ polyoxometallate: Entdeckung und untersuchung. *Chem. Unserer Zeit*, 52:2–7, 2018. doi:[10.1002/ciuz.201800818](https://doi.org/10.1002/ciuz.201800818).
- [50] T. Rahman, E. Petrus, M. Segado, N. P. Martin, L. N. Palys, M. A. Rambaran, C. A. Ohlin, C. Bo, and M. Nyman. Predicting the Solubility of Inorganic Ion Pairs in Water. *Angew. Chem. Int. Ed.*, 61:e202117839, 2022. doi:[10.1002/anie.202117839](https://doi.org/10.1002/anie.202117839).
- [51] T. Rau, M. Koch, N. Pathmanathan, T. Isenberg, D. Weiskopf, M. Sedlmair, A. Köhn, and K. Kurzhals. Understanding collaborative learning of molecular structures in AR with eye tracking. *IEEE Comput. Graph. Appl.*, 44(6):39–51, 2024. doi:[10.1109/MCG.2024.3503903](https://doi.org/10.1109/MCG.2024.3503903).
- [52] S. D. Rihm, F. Saluz, A. Kondinski, J. Bai, P. W. V. Butler, S. Mosbach, J. Akroyd, and M. Kraft. Extraction of chemical synthesis information using The World Avatar. *Digital Discovery*, 2025. doi:[10.1039/D5DD00183H](https://doi.org/10.1039/D5DD00183H).
- [53] S. D. Rihm, D. N. Tran, A. Kondinski, and et al. Natural language access point to digital metal–organic polyhedra chemistry in the world avatar. *Data-Centric Eng.*, 6:e22, 2025. doi:[10.1017/dce.2025.12](https://doi.org/10.1017/dce.2025.12).
- [54] A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein, and R. Q. Snurr. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter*, 4(5):1578–1597, 2021. ISSN 2590-2385. doi:<https://doi.org/10.1016/j.matt.2021.02.015>.
- [55] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. P. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE Trans. Vis. Comput. Graph.*, 20(12):1604–1613, 2014. doi:[10.1109/TVCG.2014.2346481](https://doi.org/10.1109/TVCG.2014.2346481).
- [56] D. Salley, J. S. Manzano, P. J. Kitson, and L. Cronin. Robotic Modules for the Programmable Computation of Molecules and Materials. *ACS Cent. Sci.*, 9:1525–1537, 2023.
- [57] K. Schwab. The fourth industrial revolution: What it means, how to respond. In *Handbook of Research on Strategic Leadership in the Fourth Industrial Revolution*, chapter 2, pages 29–34. Edward Elgar Publishing, 2024. doi:[10.4337/9781802208818.00008](https://doi.org/10.4337/9781802208818.00008).

- [58] S. L. Skjærvø, A. S. Anker, M. C. Wied, E. T. Kjær, M. Juelsholt, T. L. Christiansen, and K. M. O. Jensen. Atomic structural changes in the formation of transition metal tungstates: the role of polyoxometalate structures in material crystallization. *Chem. Sci.*, 14:4806–4816, 2023. doi:10.1039/D3SC00426K.
- [59] P. Strömert, J. Hunold, A. Castro, S. Neumann, and O. Koepler. Ontologies4Chem: the landscape of ontologies in chemistry. *Pure and Applied Chemistry*, 94(6):605–622, 2022. doi:10.1515/pac-2021-2007.
- [60] W. J. Sturm, T. Schaefer, T. Schreck, A. Holzinger, and T. Ullrich. Extending the Scaffold Hunter Visualization Toolkit with Interactive Heatmaps. In *Computer Graphics & Visual Computing (CGVC 2015)*, pages 78–84, Goslar, Germany, 2015. Eurographics Association. doi:10.2312/cgvc.20151247.
- [61] A. Tatu, F. Maaß, I. Faerber, E. Bertini, T. Schreck, T. Seidl, and D. A. Keim. Subspace Search and Visualization to Make Sense of Alternative Clusterings in High-Dimensional Data. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 63–72, 2012. doi:10.1109/VAST.2012.6400488.
- [62] L. Vilà-Nadal. Pomzites: A roadmap for inverse design in metal oxide chemistry. *Int. J. Quantum Chem.*, 121(5):e26493, 2021. doi:10.1002/qua.26493.
- [63] S.-S. Wang and G.-Y. Yang. Recent advances in polyoxometalate-catalyzed reactions. *Chem. Rev.*, 115(11):4893–4962, 2015. doi:10.1021/cr500390v.
- [64] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, and et al. The fair guiding principles for scientific data management and stewardship. *Sci. Data*, 3:160018, 2016. doi:10.1038/sdata.2016.18.
- [65] B. Zhang, M. Zhao, Y. Qi, R. Tian, B. B. Carter, H. Zou, C. Zhang, and C. Wang. The intrinsic enzyme activities of the classic polyoxometalates. *Sci. Rep.*, 9:14832, 2019. doi:10.1038/s41598-019-50539-9.
- [66] G. Zhang, Z.-Y. Xiong, Y. Gong, Z. Zhu, Z. Lv, Y. Wang, J.-Q. Yang, X. Xing, Z.-P. Wang, J. Qin, Y. Zhou, and S.-T. Han. Polyoxometalate Accelerated Cationic Migration for Reservoir Computing. *Adv. Funct. Mater.*, 32:2204721, 2022. doi:10.1002/adfm.202204721.
- [67] W. Zhang, Q. Liu, and H. Chen. Revealing catalyst genes for co2 activation via interpretable machine learning. *Nat. Catal.*, 4:345–352, 2021. doi:10.1038/s41929-021-00652-0.
- [68] Z. Zheng, Z. Rong, N. Rampal, C. Borgs, J. T. Chayes, and O. M. Yaghi. A gpt-4 reticular chemist for guiding mof discovery. *Angew. Chem. Int. Ed.*, 62(46):e202311983, 2023. doi:10.1002/anie.202311983.
- [69] Z. Zheng, Z. He, O. Khattab, N. Rampal, M. A. Zaharia, C. Borgs, J. T. Chayes, and O. M. Yaghi. Image and data mining in reticular chemistry powered by GPT-4V. *Digit. Discov.*, 3(3):491–501, 2024. doi:10.1039/D3DD00239J.