Oxygen on Mars? A task for the "machine chemist"!

Rolf Schmid[a] and Xin Xiong[b]

[a]Bio4Business, Stuttgart

[b]NMI Reutlingen

Space travel and space research are highly valued in China and are also giving new impetus to chemical research. A recent publication deals with the question of how an autonomous "machine chemist" could produce catalysts from Mars ores that would produce oxygen from Mars water on the red planet. In computer science terms, this would require a Category 4 self-driving laboratory: it would be capable of design-build-test-learn cycles (DBTL) thanks to autonomous AI-supported experimental planning and a robotic peripheral for synthesizing and analyzing the products.

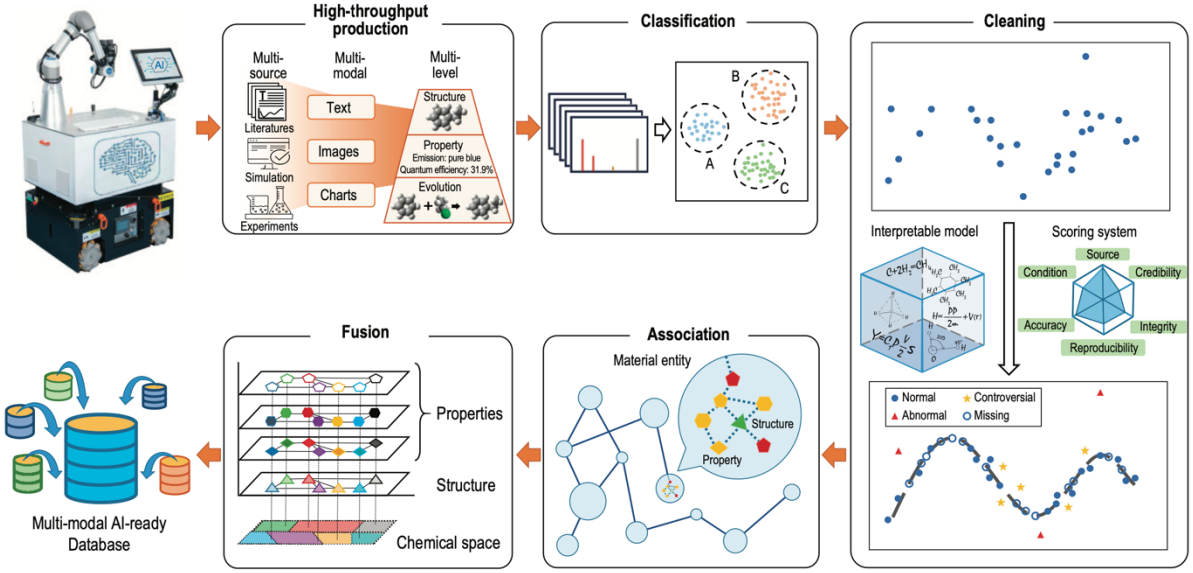
A proposed solution comes from the working groups led by Li Zhenyu and Jiang Jun at the Key Laboratory of Precision Intelligent Chemistry at the University of Science and Technology of China in Hefei, China's elite university under the Chinese Academy of Sciences. There, a robot named Xiaolai (小来, "small future") with artificial intelligence (AI) was trained to read huge amounts of chemical literature, process this knowledge into synthesis proposals, and use it to carry out synthesis, characterization, and performance tests at 14 workstations in parallel mode. After 6 weeks of work using data from around 50,000 scientific publications, Xiaolai had autonomously learned to use her robotic peripherals to produce a suitable catalyst from the minerals of the Martian rock. In an electrochemical workstation still on Earth, she then developed oxygen from water for a week under the simulated temperature conditions of the Martian surface without any loss of activity.

1. Xiaolai's design-learn function.

Like many other AI systems for chemical research, Xiaolai also draws on digitized chemical literature. However, much of the data is difficult to use to train an AI because it comes from different sources and is not quality-assured. The Hefei group is therefore building a "digital twin" of the specialist literature with standardized data using a combination of large-scale data mining and powerful computer simulations. In doing so, new approaches are being taken in the collection, classification and calibration of the data.

Data collection. First, data is extracted from texts using speech processing and image recognition. However, it is often not consistent and comparable and therefore needs to be annotated. The researchers in Hefei have developed a very fast, unsupervised method for this: a syntactic distance analysis (SDA) that uses word vectors, syntax tree and syntax distance analyses. It only needs a few keywords to mine chemical structures, functions, properties and operations.

Classification. The researchers in Hefei rely on spectra as universal, comparable, theoretically calculable and experimentally measurable descriptors for clustering the structure and properties of all chemical compounds. They use the independently developed tensor network DataNet, which predicts UV-Vis spectra with 92% accuracy and IR, Raman and NMR spectra with over 99% accuracy, 1000 – 100,000 times faster than quantum chemical predictions.



PERMISSION CC BY 4.0 DEED

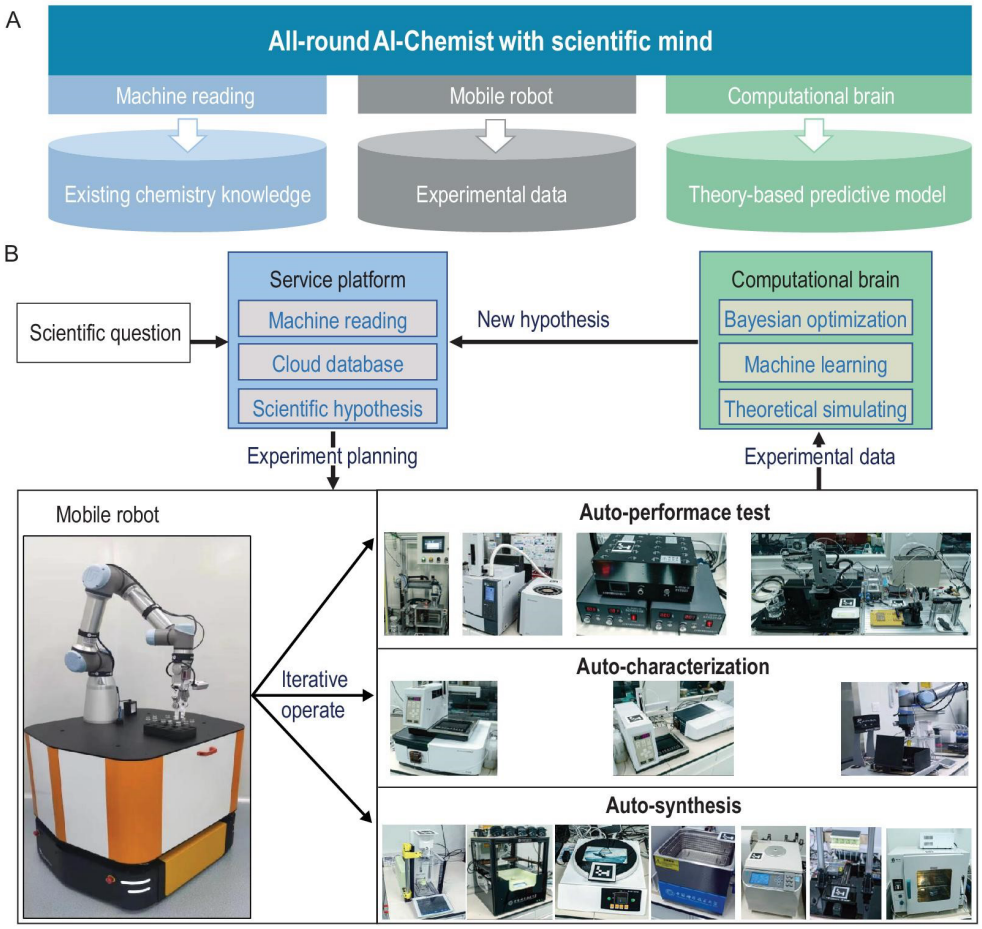
Calibration. Chemistry-based molecular graphs (CIMG) are used to describe chemical reactions in order to explore synthesis options. They have the following features: a) the chemical shifts from NMR spectra for the vertex of a reaction, b) bond dissociation energies for its edges and c) information about solvent and catalyst as global features. For the retrosynthesis of any compound, a molecular graph of the product is generated and used by a graph neural network model (GNN) to select reaction templates for the retrosynthesis. Two further GNN models with molecular graphs of the reactants are used to find suitable catalysts and solvents. Finally, a fourth GNN model checks the plausibility of the proposed reaction. When training these models, reaction vectors are obtained for each molecule. They contain the chemical knowledge about the reactivity

and accelerate the Monte Carlo tree search (MCTS) when planning multi-step retrosyntheses by categorizing the selected molecules and reactions. This molecular graph-based approach can quickly and efficiently predict complete synthesis pathways, suitable catalysts and solvents.

2. Xiaolai's build test function

A robot with peripherals performs the chemical reactions and provides the data for AI-driven optimization of the synthesis. A robotic arm with six degrees of freedom is mounted on an omnidirectional mobile platform, capable of supporting up to 200 kg, and locates samples with an accuracy of 10 mm. With an integrated dual lidar-based mapping and localization system, it accesses 14 workstations. The laboratory, which houses robots and equipment, is divided into three areas for synthesis, characterization and performance testing. The liquid dispensing workstation operates with an accuracy of 3 μL, and the solid dispensing workstation operates with an accuracy of 0.1 mg. The magnetic stirring and ultrasonic workstations operate with an accuracy of one millisecond. The cleaning workstation allows centrifugation and liquid extraction. After the synthesis is complete, the product is characterized and tested for performance. Workstations for UV-Vis, fluorescence, Raman spectroscopy and cyclic voltammetry are available. Other workstations are used for photocatalysis and gas chromatography. A sealing station allows reactions under pressure or in a vacuum. For high-throughput and multitasking experiments, there is a large and scalable loading and delivery platform. All instruments are designed to operate with vials of the same type to enable passage through multiple workstations.

3. Xiaolai - a Category 47 Self-Driving Laboratory



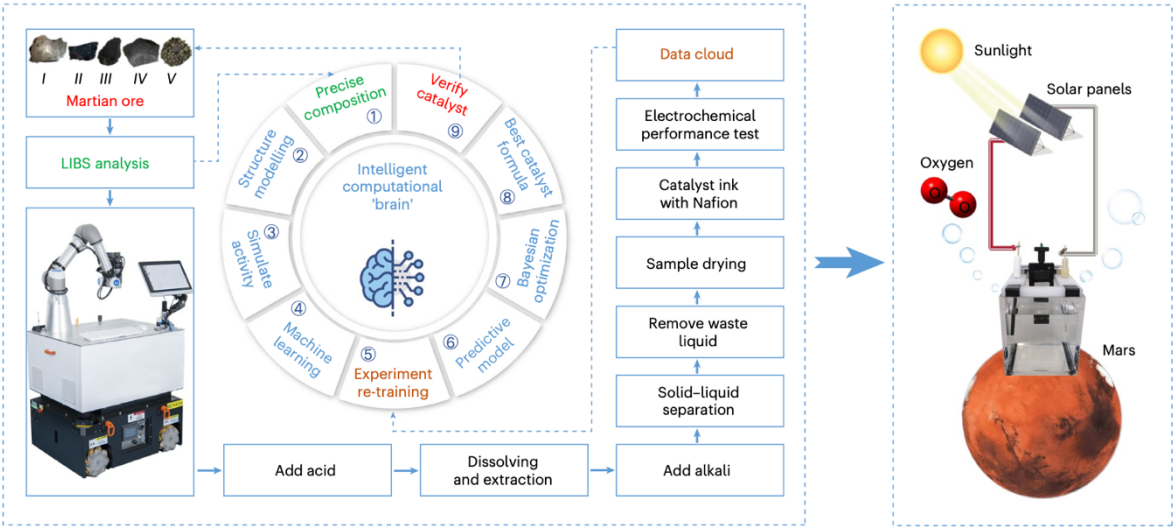
PERMISSION CC BY 4.0 DEED

When Xiaolai receives a new task, its user-friendly graphical interface suggests a solution that it has developed from existing knowledge by machine reading large amounts of literature. It extracts common patterns and converts them into a uniform, scalable, structurally unambiguous and multimodally oriented data format that contains material structure, properties and reaction characteristics and is suitable as input for AI models. Combined with theoretical simulation and machine learning models, it is a digital twin of a chemist and can predict and optimize properties and syntheses of materials. Xiaolai can access data to design experimental plans and use a web-based function to distribute tasks that are carried out by the mobile robot at the various workstations. However, Xiaolai can not only capture existing chemical knowledge, but also find a globally optimal solution by conducting theoretical simulations, training machine learning models and performing Bayesian optimizations. It can independently propose new hypotheses and experimental plans and thus independently carry out the next round of chemical experiments.

4. Results

The capabilities of Xiaolai have been described in several publications. For example, the system independently evaluated 15,979 scientific papers to synthesize a biocompatible luminophore with aggregation-induced emission properties, optimized the hydrogenation step of a HxMoO3 photocatalyst, and developed chiral films with high chiro-optical activity.

To find a catalyst for obtaining oxygen on Mars, Xiaolai analyzed about 50,000 papers on the composition of Martian meteorites and water-splitting catalysts of similar composition and designed a base material optimized according to the DBTL cycle. The entire manufacturing process, including the pretreatment of the Martian ore, catalyst synthesis, its characterization and optimization, was carried out without human intervention. After six weeks, the best formula was found: the catalyst material consisted of Mn, Fe, Ni, Mg, Al and Ca. In Xiaolai's electrochemical workstation, it formed oxygen from water for 550,000 s (153 hours) at an overpotential of 445.1 mV and a current density of 10 mA cm−2. And it did so without any loss of performance at −37 °C, the simulated temperature conditions of the Martian surface2.



5. What is the international competition doing?

Countless publications describe the data mining of chemical literature as a supporting or autonomous guide for retrosyntheses. More recent is the integration of this concept into a “self-driving lab”, often also called a material acceleration platform (MAP). Modular in design, these stations autonomously carry out syntheses, independently analyze and interpret the materials obtained and iterate this process using machine learning. Bioresearch has long been studying “biofoundries”, in which a robot-controlled workstation synthesizes DNA strands with high throughput in order to carry out enzyme or genome design. This produces technical enzymes or “smart cells” (microorganisms,

plants) for biotechnological processes. The pharmaceutical industry uses similar concepts in drug research – one example is Novartis' Micro-Cycle project.

Learning robots with their own robot peripherals for the synthesis of solids or organic active substances are, however, still quite new. Table 1 contains a selection of players. In Germany, the Helmholtz Center in Erlangen operates a self-controlled autonomous system for the development of new photovoltaic technologies. The Helmholtz Center Ulm is using self-controlled systems to research new battery materials. And at the Federal Institute for Materials Research and Testing BAM in Berlin, a MAP is currently being set up that deals with the development of nano and advanced materials, materials for the energy transition and sustainable building materials. It is intended to accompany material developments down to the component and device level.

Table 1: Selection of players in the field of autonomous chemical workstations

PI	**Location**	**Features**	**Application examples**
Lee Cronin Robot	U. of Glasgow	workstation with "Chempiler" software	"Chemputer"
Andrew I Cooper	U. of Liverpool	AI-based robot workstation	Photocatalytically active compounds
Klavs F Jensen	MIT	Autonomous, AI-based robot workstation	Dyes, active ingredients, electrochemical reactions
Yan Zeng and Gerbrand Ceder	Lawrence Berkeley National Laboratory	A-Lab: Robot workstation with Google Deep Mind	Solid synthesis of inorganic powders
Alan Aspuru-Guzik	U. of Toronto, The Matter Lab and Intrepid Lab	AI-based robot workstation with ChemOS	Organic catalysis
Timothy Noel	U Amsterdam	RoboChem – autonomous synthesis platform	Photocatalysis
Curtis P. Berlinguette	U. of British Columbia, Vancouver	Autonomous workstation ADA	Film materials for coating
Youn-Suk Choi and Youngchoun Kwon	Samsung Advanced Institute of Technology	AI-based robot workstation	AI-based synthesis of organic molecules
Jens Hauch and Christoph J Brabec	Helmholtz Institute Erlangen and U.	AMADAP, a semi-automatic synthesis platform	Photocatalytically active compounds

	Erlangen-Nuremberg		
Helge Stein	Helmholtz Institute Ulm, BIG MAP	BIG MAP Battery Interface Genome Materials Acceleration Platform	Battery research
Bastian Rühle and Özlem Özcan-Sandikcioglu	Federal Institute for Materials Research and Testing, Berlin	Autonomous synthesis platform	Nano- and advanced materials, materials for the energy transition and sustainable building materials

With its AI robot Xiaolai and its extensive robot peripherals, the working group in Hefei is one of the leading global developments. What does Jiang Jun, who was also head of his university's youth class for a year in his spare time, say about this? "The all-round AI chemist has created a team that knows something about chemistry, computers, hardware design and software architecture. The cross-training model of our university's youth class aims to not immediately divide students into major subjects. They first learn knowledge aimlessly, just like ChatGPT."

However, the machine chemists could soon face competition from biology when it comes to the production of oxygen on Mars: another Chinese team recently proposed colonizing the surface of Mars with a desert moss that is extremely resistant to temperature and radiation and could produce oxygen through photosynthesis.

We thank Prof. Dr. Jürgen Pleiss, University of Stuttgart, and Dr. Bastian Rühle, Federal Institute for Materials Research and Testing, Berlin, for reading the manuscript and for valuable comments.

[1] https://en.ustc.edu.cn

[1] *Q. Zhu, Y. Huang, D. Zhou, L. Zhao, L. Guo, R. Yang, Z. Sun, M. Luo, F. Zhang, H. Xiao, X. Tang, X. Zhang, T. Song, X. Li, B. Chong, J. Zhou, Y. Zhang, B. Zhang, J. Cao, G. Zhang, S. Wang, G. Ye, W. Zhang, H. Zhao, S. Cong, H. Li, L.-L. Ling, Z. Zhang, W. Shang, J. Jiang and Y. Luo: Automated synthesis of oxygen-producing catalysts from Martian meteorites by a robotic AI chemist. In:* Nature Synthesis. 3, (2023), S. 319-328 (doi: 10.1038/s44160-023-00424-1)

[3] *S. Feng, A. R. Cai, Y. Wang, B. Zhang, Q. Qiao, C. Chen, S. Wang, J. Jiang: A robotic AI-Chemist system for multi-modal AI-ready database. In: National Science Review. 10, (2023), (doi: 10.1093/nsr/nwad332)*

[1] *B. C. Zhang, H. Y. Xiao, G. L. Ye, Z. K. Song, T. T. Han, E. Sharman, M. Luo, A. Y. Cheng, Q. Zhu, H. T. Zhao, G. Q. Zhang, S. Wang, and J. Jiang. Label-Free Data Mining of Scientific Literature by Unsupervised Syntactic Distance Analysis. In: The Journal of Physical Chemistry Letters. 15, (2024), S. 212-219 (doi: 10.1021/acs.jpclett.3c03345)*

[1] Y. Y. Chong, Y. Y. Huo, S. Jiang, X. J. Wang, B. C. Zhang, T. F. Liu, X. Chen, T. T. Han, P. Smith, S. Wang, and J. Jiang. Machine learning of spectra-property relationship for imperfect and small chemistry data. In: PNAS. 120, (2023), (doi: 10.1073/pnas.2220789120)

[1] Z. Zou, Y. Zhang, L. Liang, M. Z. Wei, J. C. Leng, J. Jiang, Y. Luo, and W. Hu. A deep learning model for predicting selected organic molecular spectra. In: Nature Computer Science. 3, (2023) S. 957–964 (doi: 10.1038/s43588-023-00550-y)
[1] Q. Zhu, F. Zhang, Y. Huang, H. Y. Xiao, L. Y. Zhao, X. C. Zhang, T. Song, X. S. Tang, X. Li, G. He, B. C. Chong, J. Y. Zhou, Y. H. Zhang, B. C. Zhang, J. Q. Cao, M. Luo, S. Wang, G. L. Ye, W. J. Zhang, X. Chen, S. Cong, D. L. Zhou, H. R Li, J. L. Li, G. Zou, W. W. Shang, J. Jiang, and Y. Luo, An all-round AI-Chemist with a scientific mind, In: National Science Review, 9, (2022), nwac190, (doi.org/10.1093/nsr/nwac190)
[1] Xie, Y., Feng, S., Deng, L. et al. Inverse design of chiral functional films by a robotic AI-guided system. In: Nature Communication, 14, (2023), (doi: 10.1038/s41467-023-41951-x)
[1] https://de.wikipedia.org/wiki/Marsmeteorit
[1] S. Back, A. Aspuru-Guzik, M. Ceriotti, et al., Accelerated chemical science with AI. In: Digital Discovery, 3, (2023) (doi: 10.1039/d3dd00213f)
[1] Nachrichten aus der Chemie | 71 | Mai 2023 |
[1] C. Brocklehurst, E. Altmann, C. Bon, et al., MicroCycle: An Integrated and Automated Platform to Accelerate Drug Discovery. In: Journal of Medicinal Chemistry, 67, (2024), S. 2118-2128, (doi: 10.1021/acs.jmedchem.3c02029)
[1] http://www.hfnl.ustc.edu.cn/detail?id=20813
[1]X. S. Li, W. W. Bai, Q. L. Yang, et al., The extremotolerant desert moss Syntrichia caninervis is a promising pioneer plant for colonizing extraterrestrial environments. In: The Innovation, 5, (2024), (doi: 10.1016/j.xinn.2024.100657)