

Both the renewable and nuclear energy sectors have served as catalysts for significant advancements in materials chemistry, driving the development of new materials that enhance the efficiency, sustainability, and safety of energy technologies. This symbiotic relationship highlights the critical role of materials chemistry in powering the future with a blend of sustainable and reliable energy sources, showcasing the field's extensive contributions to addressing global energy challenges.

These advancements are part of a broader trend where materials chemistry is pivotal in tackling the energy challenges of the twenty-first century. The discipline not only provides new materials but also offers insights into the mechanisms at play within these materials, enabling the design of systems that are more efficient, sustainable, and environmentally friendly. Additionally, the intersection of materials chemistry with other industries has led to a fruitful exchange of ideas and technologies. For instance, developments in photovoltaic materials influence the automotive industry by enabling the creation of lighter, more energy-efficient vehicles. Similarly, advancements in materials for energy storage have significant implications for both the grid and transportation sectors, potentially revolutionizing energy consumption patterns.

In summary, while new energy forms have not fundamentally changed the theoretical or experimental foundations of materials science research, they have greatly expanded the field's scope and impact. The evolution of materials chemistry in response to new energy challenges exemplifies the dynamic nature of scientific inquiry, where advancements in one domain catalyze progress across multiple areas. This interplay between materials chemistry and new energy technologies underscores the field's crucial role in fostering a sustainable and technologically advanced future.

1.3 Data

The advancement of science and technology has significantly increased the precision of experimental testing, fundamentally improving our understanding of the relationships between structure and properties in materials chemistry. Enhanced experimental techniques, such as high-resolution microscopy and sophisticated spectroscopy, allow for detailed observations of materials at the atomic level. These insights are crucial for identifying how structural nuances influence material properties.

Furthermore, the integration of computational models with experimental data accelerates the exploration of new materials, enabling predictions about how changes in structure affect properties before physical experiments are conducted. Techniques like chemical vapor deposition (CVD) and atomic layer deposition (ALD) exemplify the precise control over experimental conditions necessary for synthesizing advanced materials with specific desired properties.

This improvement in experimental precision not only propels basic research but also drives technological innovations, such as more efficient solar panels and batteries with higher energy densities. As experimental methods continue to evolve, they

promise to unveil new phenomena, enhancing our understanding of materials and opening up further possibilities for innovation. In essence, the progress in experimental accuracy enriches our material knowledge, catalyzing advancements across both scientific and technological landscapes.

From the era of steam engines to the electrical age, humanity has undergone a transformation in the way we harness energy, thereby accelerating the processes of material production and synthesis. This has naturally accumulated a vast amount of knowledge, transmitted through education across generations. However, due to various reasons, it is inevitable that knowledge will be lost or even inaccurately transmitted during the educational process. For example, the question of what a “dragon” looks like in Chinese mythology is difficult to answer definitively. Therefore, with the widespread adoption of computer hardware technology, transforming knowledge into data and storing it in hard drives can preserve it for many years. The key concepts of “data” and “many years” lead us to the fundamental characteristic of true technology: the ability to traverse time or space. We can express this in more complex terms using the language of physics: the implications of Noether’s theorem (Noether 1918).

To illustrate this with a familiar example, consider the camera technology we are now very familiar with, which allows us to see photographs and images from decades ago. Similarly, telephone communication technology enables us to engage in real-time communication with friends thousands of kilometers away or even in space stations, disregarding spatial distance. Nowadays, camera and telephone technologies can be seamlessly integrated into a single smartphone, which is why companies excelling in this field, such as Apple, have become among the world’s most profitable companies – a fact that should come as no surprise.

Let’s consider the development of materials chemistry and the potential challenges it may face using the logic of time and space traversal. Firstly, chemicals themselves can be a part of smartphones, which is enough to demonstrate the importance of materials chemistry. This conclusion is correct, but the logic may not be. Currently, we generally believe that the emergence of products like smartphones has gone through a progression from non-smart to smart phones. The competitiveness of these products is not solely driven by their material composition. In simpler terms, the logic behind smartphone development is not primarily dominated by materials science.

Take another example, the pharmaceutical industry. The significance of high-quality drugs does not need overstating, and compared to the smartphone industry, the connection to chemistry is more intricate here. In a specific drug development process, numerous chemical molecules are tested based on demand, which is a complex and enduring process. Each chemical molecule in any testing stage may be proven unable to progress to the next stage. There are many blind spots in clinical stages, and selection can only be made based on experimental results. Therefore, drug development is a high-investment, high-risk industry. Thus, whether in the smartphone or pharmaceutical industry, chemical materials are objects of selection rather than having the power of active choice. In simple terms, many industries require the properties of chemical materials, but these properties cannot be mapped

directly to structure, necessitating a lot of experimentation. This extensive experimentation itself consumes a lot of time and resources, and even after numerous attempts, the expected results may not be achieved. Meeting the expected results at one stage does not necessarily mean meeting the expected results at the next stage, leading to wasted efforts. Understanding this, you might understand why materials chemistry is not a popular major in universities, especially when a large number of repetitive experiments currently rely on human labor, with low success rates after time investments. Of course, there are also factors such as high personal safety risks and high degree of repetition.

From the described phenomena above, two main problems arise:

1. The ambiguous relationship between structure and performance.
2. Lengthy testing cycles for experimental data.

We see that these are still problems related to time and space. The corresponding solutions are, then:

1. Clarifying the complex relationship between structure and performance.
2. Shortening the testing cycle.

These solutions might seem like they're saying everything and nothing at the same time. Indeed, many problems have similar answers. The answer to the first problem is knowledge, while the answer to the second problem is execution. Interpreted through ancient Chinese wisdom, this is "unity of knowledge and action."

Let's first explore the realm of knowledge. We've elaborated on the journey from the era of steam engines to the current electrical and new energy era over the past 200 years. Throughout this process, the discipline of materials chemistry has been present and has accumulated a vast amount of knowledge and technology. Objectively, it has also promoted the development of other disciplines. However, the primary contributors and holders of this knowledge and technology have been the traditional developed countries such as Europe, North America, and Japan. Most of the powerful materials chemistry giants undoubtedly hail from these countries.

Now, let's briefly examine the situation of materials chemistry in China. Since joining the World Trade Organization in 1999, China's chemical industry has undoubtedly become one of the most important components of the global chemical industry. Chinese universities and research institutes have surpassed the United States in publishing academic papers in the field of chemical materials, becoming the world leader. It took less than 30 years for China to catch up from a relatively backward state, which could not have been achieved without mutual learning and communication with countries worldwide. The most significant aspect we observe is the tremendous change brought about by the transmission of knowledge. We can roughly estimate that China has learned and mastered approximately 200 years' worth of knowledge in just 30 years, leading to enormous economic and academic growth.

Let's carefully contemplate the two numbers and the logic behind them. The process of these 30 years is driven by continuous human effort, whether it's physical laborers or academic workers contributing. The most typical data point

is the significant increase in urbanization rate and the expansion of universities, both in terms of the number of students and research personnel. In line with our previous discussion, this is a chemically driven process. However, the essence of the knowledge gained over these 200 years is essentially stock knowledge. In terms of incremental knowledge, cutting-edge biotechnological materials technology still needs to be imported from Western countries. Here, we would not go through the international geopolitical and economic issues. If we merely consider China's materials chemistry industry as a model, we can summarize its characteristics: acquiring a vast amount of knowledge in a short period and naturally predicting its future demand, which is to acquire even more knowledge in a shorter time. It's essential to note that here we are referring to "knowledge," not just "technology."

Let's take a look around. In today's electrified age, a large number of factories, including chemical plants, produce various chemical industrial products in large quantities through automation technology. One reason for this mass production is that mechanization and automation have replaced manual operations with standardized procedures, thereby significantly increasing the number of finished products per unit time and reducing costs. Robots and automation technology are widely used in the industrial sector, enhancing production efficiency and product quality. However, their application in chemical, materials, and biological laboratories is relatively limited. Why is that?

The reason lies in the stark differences between technological implementation in industry and knowledge realization in academia, in terms of quantity and process requirements. One involves the systematic digestion of existing knowledge, while the other entails free exploration for incremental transformation. Thus, even though chemical plants around us started automating and digitizing much earlier than 50 years ago, and despite having more and better-quality experimental testing equipment, the layout of chemical materials laboratories has not changed significantly. As shown in the images below, on the left is a picture of a chemistry laboratory at Beijing Institute of Technology in 1960, and on the right is a picture of a chemistry laboratory at a certain university in 2024. Both images show rows of experimental workbenches, cabinets underneath, and many chemical reagents and drugs on the workbenches. Despite spanning 64 years, there is no significant difference between the two images. We still rely on our graduate students and postdocs to operate chemical containers and obtain experimental data, and the entire process of chemical experimentation has not achieved automated and intelligent assembly-line-style operation (Figure 1.4).

Indeed, we can argue that the complexity and variability of laboratory work, along with the need for flexibility and specialized knowledge, make it challenging to achieve automation. However, from a solution perspective, there is not much difference compared to solving other physical chemistry problems. We start by addressing one point, then move on to handling the relationships between this point and others, and gradually address the relationships between more points. Although the reagent bottles used in chemical experiments are standardized – for example, in my laboratory, we only have three types: 10, 100, and 500 ml bottles – we cannot demand standardization when it comes to the actions of our students. For

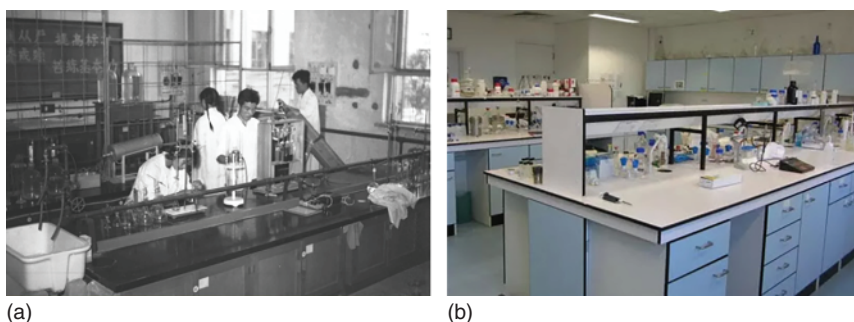


Figure 1.4 The chemistry lab in (a) 1960s (SinaEducation 2010) and (b) 2024.

instance, we cannot insist that they must assemble a 100 ml reagent bottle within one minute, and that the turbulence bubbles in the solution must be uniform. While the latter may be difficult to achieve, the former could be easily realized through automation technology. However, it may not seem necessary to limit the assembly time to one minute until we realize that the time for reagent preparation may have subtle connections with the quality of the final product.

There is another crucial logical issue here. In the history of chemical materials development, many significant discoveries have occurred due to accidentally using the wrong reactants or incorrect amounts of reactants. For example, the discovery of conducting polymers, which began with Hideki Shirakawa's accidental discovery in 1971 (Shirakawa and Ikeda 1971). At the time, he was supervising a group of South Korean exchange students conducting experiments on the synthesis of polyacetylene. Perhaps due to a language barrier, the students mistakenly increased the amount of the designated catalyst, the Ziegler–Natta catalyst, by thousands of times. As a result, they produced a silver-colored polyacetylene film, overcoming the challenge of synthesizing highly crystalline polyacetylene. The lightweight structure obtained provided an excellent foundation for future doping to enhance electrical conductivity. If Shirakawa had used a machine to precisely conduct the experiment, such an important discovery might not have been made. Although this is a relatively specific example, it is certainly not an isolated case. Historically, laboratory practitioners seem to not prefer automated equipment and consider experimental operations as an integral part of the experiment itself. There are two specific issues here.

Firstly, many researchers do not handle multiple experimental tasks and requirements simultaneously. In the same period, many experimenters often only need to focus on a specific project. This may be determined by the project's budget, timeline, and the current research paradigm. Conversely, pharmaceutical companies have different practices. Over 30 years ago, many pipetting workstations were already in operation. Although these workstations generate vast amounts of research data every day, when I deployed such a workstation at a Chinese university in 2023, I found that the doctoral students there did not frequently use the machine. One reason is that they could not handle so many samples at once.

We have learned about the significant differences between university laboratories and corporate laboratories, as well as the reasons why it is more challenging for university laboratories to adopt intelligent and automated equipment: the conflict between academic credit and task division. In a corporate setting, each position has very detailed task division, so there are clear application scenarios for accelerating processes through automation equipment. However, university laboratories are different. Academic credit for research achievements is a crucial indicator, so it is not feasible to distribute tasks or outsource them through assembly-line-style operations. Researchers theoretically need to be familiar with all aspects of the research process, especially many sequential steps. Researchers need to understand detailed information about all aspects to design and correct the next experimental operation. Current experimental automation equipment focuses on the needs of pharmaceutical companies and other discrete experiments within enterprises but does not provide intelligent equipment with a full-process information architecture. Therefore, for specific chemical materials researchers, the use of local automation acceleration equipment, including pipetting workstations, does not significantly promote research efficiency. Therefore, the main reason why the current paradigm of chemical material science research, mainly led by universities and research institutes, has not achieved a paradigm shift is that there is no automation stage, and it must directly enter intelligence.

I mentioned earlier that in the future, whether in China or the world, the chemical materials industry or any other industry will need to master more knowledge in a shorter time. The above paragraph explains why “intelligent” equipment, rather than “automatic” equipment, is the solution for “shorter time.” Next, we need to explain what the solution for “more knowledge” is. Suppose a laboratory researcher synthesizes a chemical molecule that contributes to the treatment of a certain disease. Is this molecule knowledge? Of course, it is, but it is far from enough. The full process information including reaction conditions, synthesis pathways, and various experimental details involving this molecule is more critical knowledge, and the amount of knowledge is closely related to the amount of information. Here, we need to mention the difference between patents and papers. In order to protect the intellectual property related to the knowledge, a patent related to molecule synthesis often discloses many process details, which are protected commercially, but still promote knowledge dissemination. Authors and readers of academic papers often focus more on the conclusions themselves and may not deliberately detail the implementation process because they care more about personal academic credit, and academic papers themselves are not a form of protectable intellectual property. Of course, there are many other problems in the academic paper field, such as unreliable conclusions, poor reproducibility, repetitive research, academic integrity, and so on, which can seriously hinder or even mislead knowledge dissemination. There are countless similar events, and dealing with these problems is very difficult. For example, an American scientist misled the work of 100 research groups in China for 10 years with a false paper, and there is currently no organization that can handle such incidents. In comparison, patents are much simpler, and problems can be solved through domestic or international legal means (Figure 1.5)

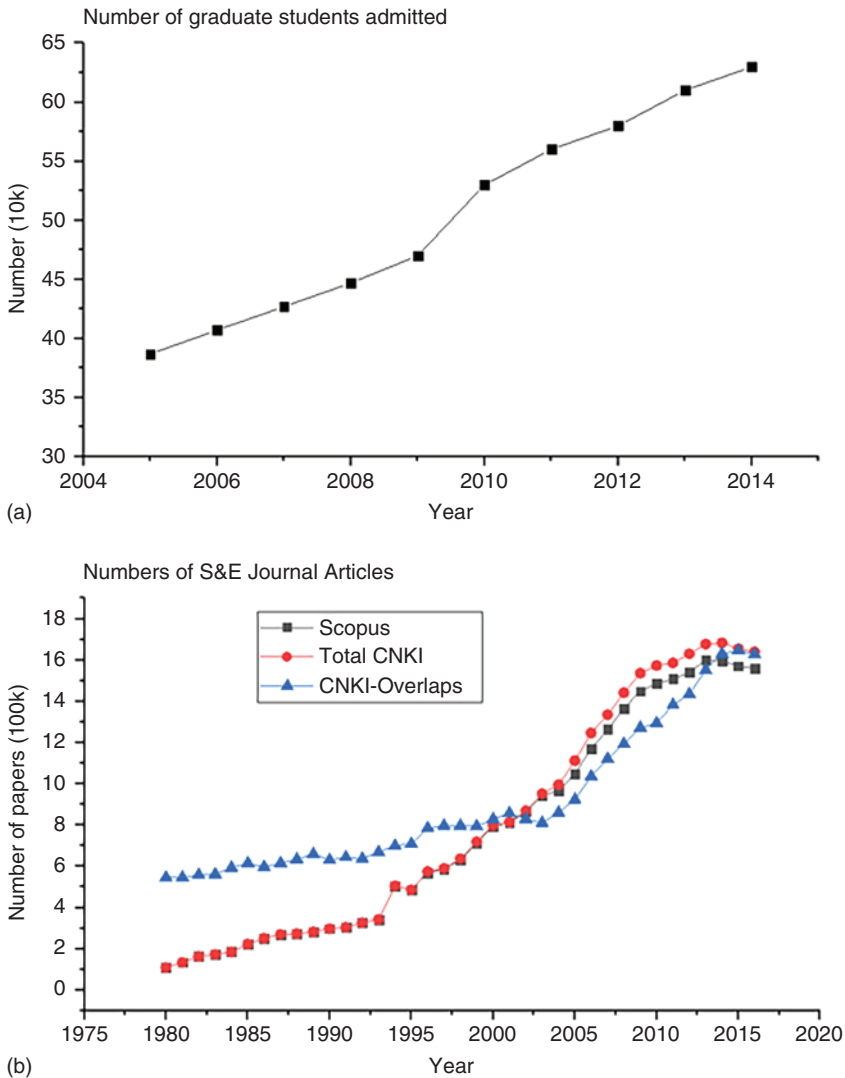


Figure 1.5 (a) The number of graduate students in China has grown since 2004 by about 80% by 2014 (ChinaEducationOnline 2015). (b) The numbers of S&E Journal Articles in Scopus and in CNKI, 1980–2016 (VoxChina 2018).

Let's discuss the issue of the rate of knowledge generation using academic papers as an example. A report (Miyazaki 2023) by *Nature* journal in 2023 stated that China had surpassed the United States in the number of high-quality journals indexed in the Nature Index, with the number of journals in the field of chemical materials surpassing even earlier. There are certainly many reasons why China has achieved this surpassing. Let's take a look at the two charts above. We can observe that the number of graduate students enrolled in China has increased from approximately 360,000 in 2005 to about 640,000 in 2014, an increase of about

80% (ChinaEducationOnline 2015). Meanwhile, the number of scientific and technological papers has increased from about 800,000 in 2005 to about 1.6 million in 2014, a growth of about 100% (VoxChina 2018). We notice that these two numbers are very close, indicating that the increase in the number of papers in China cannot be separated from the expansion of the number of graduate students, including in the field of chemical materials science. If we assume that, on average, an academic paper in the field of chemical materials represents the synthesis or discovery of a new material, then is the rate of discovering new materials meaningful?

Actually, it is not meaningful because it is still too slow! We can do a simple and rough calculation to estimate how many chemical experiments we can conduct. Figure 1.6 summarizes and classifies the total ~2 million papers published in the field of materials science in 2022 through our KnowledgeWorks (Yanheng et al. 2024), divided into 15 categories based on application scenarios, including Energy Storage and Conversion, Catalysis, Environmental Applications, and other familiar categories. Each application scenario can be achieved through a maximum of eight experimental methods. Let's assume there are a total of 1 million reagents, 10 reaction steps, 100 solvents, and up to 4 solvent mixtures, as well as approximately 100 reaction conditions for each of the 10 reaction steps. Calculating this, we find:

$$15 \text{ Applications} \times 8 \text{ Methods} \times (10^6 \text{ Reagents})^4 \times (10^3 \text{ Conditions})^{10} > 10^{100}$$

This means that exploring all known chemical reactions would require a dataset of one Googol (10^{100}), while the amount of data currently explored by humanity is probably less than 10^{10} . Perhaps under the current research paradigm, we can still make many significant discoveries, but the overall conquest still seems far out of reach (Figure 1.6).

This is the biggest challenge in the field of chemical materials science. Currently, it is still driven by human chemistry, much like agricultural societies, where the production of agricultural products is almost entirely linearly dependent on the number of laborers, and a country's overall strength is closely related to its population. Therefore, we often see in some countries that pursue scientific research outcomes, many research teams with hundreds of members, most of whom are graduate students. If we liken the discovery or synthesis of new chemical substances to a treasure hunt

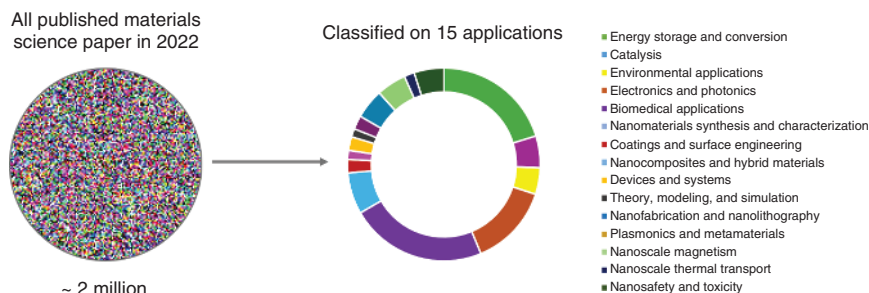


Figure 1.6 There have been ~2 million papers published in materials science since 2022. The papers can be classified according to applications into 15 categories, as shown in the figure. They can also be classified to 8 methods, 10^6 reagents, 10^2 solutions, 10^3 conditions.

game, are players driven by electricity more likely to discover targets faster and more often than players driven by chemical energy? This question is actually difficult to answer because both intelligence and physical strength are indispensable in a treasure hunt game. Electricity may make physical activities easier, but this logic does not necessarily apply to intelligence. So, let's switch to a different analogy: if the same person drives a Tesla and rides a horse, which situation is more likely to discover targets faster and more often? The answer to this question is probably self-evident.

So, can using intelligent machines alone accelerate the discovery of new materials and processes? The answer is also negative. Machines only provide a linear increase in experimental operations compared to human researchers, and this linear coefficient is not particularly large. From my personal experience, it may be around 20 times. The biggest problem with linear relationships is their linear relationship with time, meaning that achieving large-scale knowledge innovation will take a very long time in the future. This is not a reasonable scientific framework; it lacks imagination. Let's imagine a completely new mathematical formula:

$$F_{n+1} = F_n + F_{n-1}$$

Now let's define $n + 1$ as tomorrow, n as today, and $n - 1$ as yesterday. Let the function F represent the amount of data in a particular industry or chemical experiment. Then this sequence is the famous Fibonacci sequence, with the first few terms being: 1, 1, 2, 3, 5, 8, 13, 21, 34, and so on. The general form is: $F_n = \frac{\phi^n - (-\phi)^{-n}}{\sqrt{5}}$, where $\phi = \frac{1+\sqrt{5}}{2}$. When n is large, F_n is proportional to ϕ^n . Finally, we see an exponential growth relationship.

At first glance, does this look very familiar? Yes, it is indeed the mathematical principle of the famous golden ratio, where the artistry of science is vividly demonstrated. Taking a closer look, it becomes even more familiar. Essentially, this is what current large language models are doing: packaging all historical data for training, starting with F_{n-1} , and then incorporating all incremental data from today onward. For text, video, and other information on the internet, due to relatively uniform data formats and the accumulation of technical expertise in this area, technologies like ChatGPT and Sora have been impressive since their inception. If it were possible to generate new chemical materials with just a prompt word like Sora, how would we need to reconstruct our databases and algorithms?

References

- Arrhenius, S. (1889). Über die Dissociationswärme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte. *Zeitschrift für Physikalische Chemie* 4 (1): 96–116.
- Ashcroft, N.W. and Mermin, N. (1976). *Solid State Physics*. Saunders Collage Publishing.
- ChinaEducationOnline (2015). The popularity of postgraduate applications slows down and the enrollment crisis begins to spread. Retrieved from <https://www.eol.cn/html/ky/report2015/a2.shtml>.

- Griffiths, D.J. and Schroeter, D.F. (2018). *Introduction to Quantum Mechanics*. Cambridge University Press.
- Miyazaki, A. (2023). New Nature Index data shows China on top for natural sciences as US leads on health sciences research. Retrieved from <https://www.natureasia.com/en/info/press-releases/detail/8966>.
- Noether, E. (1918). Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pp. 235–257.
- Röntgen, W.C. (1896). On a new kind of rays. *Science* 3 (59): 227–231.
- Shirakawa, H. and Ikeda, S. (1971). Infrared spectra of poly (acetylene). *Polymer Journal* 2 (2): 231–244.
- SinaEducation (2010). Old photos of Beijing University of Technology: Chemistry Experiment Class in the Laboratory (pictures). Retrieved from <http://edu.sina.com.cn/gaokao/2010-07-13/1131258180.shtml>.
- VoxChina (2018). China's Overwhelming Contribution to Scientific Publications. Retrieved from <https://voxchina.org/show-3-99.html>.
- XU, Y., Ye, S., Liu, R. et al. (2024). Spider Matrix: Towards Research Paper Evaluation and Innovation for Everyone. *ChemRxiv*. doi: 10.26434/chemrxiv-2024-689vs-v4.