Artificial Intelligence-driven Nanozyme Design and Development Research Progress in Data, Models, and Closed-loop Systems

Keying Ling[1], Hongying Zhu[1], Ling Lin[1], Guizhang Wang[1], Yu Li[1], Qijia Jiang[2], Zixuan Wang[2], Xinling Lai[2], Jiaqi Xiao[1], Shuang Zou[1,*]

[1]School of Packaging Engineering, Hunan University of Technology, Zhuzhou 412007, China

[2]College of Science and Technology, Hunan University of Technology, Zhuzhou 412007, China

Keying Ling and Hongying Zhu are co-first authors, *Corresponding email: 19374285193@163.com

Abstract

Nanozymes, as a class of nanomaterials with enzyme-like catalytic activity, combine the high catalytic efficiency of natural enzymes with stability, ease of modification, and low cost of nanomaterials, making them a research hotspot in the field of electrochemical sensors. This article systematically reviews the classification, catalytic mechanisms, and performance regulation methods of nanozymes, with a focus on the construction strategies and application progress of different types of nanozymes (metal-based, carbon-based, metal-organic framework-based, etc.) in electrochemical sensors. It covers multiple fields, including biomarker detection, food safety monitoring, and environmental pollutant analysis. Furthermore, the article analyzes the current bottlenecks in the practical applications of nanozyme-based electrochemical sensors, including insufficient catalytic selectivity, poor long-term stability, and difficulty in eliminating interference from actual samples. Finally, it provides an outlook on future development trends, aiming to offer references and ideas for subsequent research in this field.

Keywords

Nanozyme, Electrochemical sensor, Catalytic mechanism, Detection application, Performance regulation

Introduction

The concept of nanozymes can be traced back to 2007, when magnetic iron oxide nanoparticles were found to have inherent peroxidase-like activity. Since then, the combined advantages of nanomaterials, including “enzyme-like catalytic sites, nanoscale effects, and engineerable interfaces” have propelled nanozyme research from accidental discovery to systematic construction and functional expansion [1]. In biomedical settings, nanozymes can serve as natural enzyme substitutes in cascade reactions, reactive oxygen species regulation, and antioxidant regulation. They can also couple with external fields such as light, magnetism, and ultrasound to achieve a controllable activation strategy for integrated diagnosis and treatment [2].

However, nanozyme research has also revealed a core bottleneck associated with the spatial complexity of high-dimensional materials. The kinetic parameters of similar materials under different synthetic routes, particle sizes and morphologies, surface ligands, and reaction conditions are dispersed and difficult to compare laterally. Their activity and selectivity are often highly sensitive to microstructure and interface states, thus limiting the application of reusable design principles. The essence of this dilemma lies in the fact that the performance of nanozymes is determined by a combination of factors, including composition, structure, morphology, surface chemistry, and reaction conditions. Traditional experimental methods struggle to systematically explore the optimal combination within such a high-dimensional parameter space [3].

Over the past decade, the rational design of nanozymes has gradually shifted from “empirical rules plus single-variable optimization” to “mechanism descriptors plus high-throughput computation and experimentation”. For example, the crystal-field electron e_g occupancy number has been proposed to predict the peroxidase-like activity of spinel oxides, establishing an operable activity descriptor framework [4]. Furthermore, researchers have proposed t_{2g} occupancy as a more

general predictive descriptor in tetrahedral crystal fields. They then combined it with a second descriptor to construct a three-dimensional "volcano-type" correlation, achieving targeted optimization from descriptor to material formulation [5]. These works demonstrate that the route of "computable descriptor-predictable model-verifiable material" is taking shape, but its scaling is still limited by data volume and experimental throughput [6]. Resonating with the above-mentioned needs, the artificial intelligence infrastructure in the field of materials science entered a period of accelerated development between 2022 and 2025. Large-scale graph networks and deep learning models have shown strong generalization capabilities in crystal stability prediction and material discovery [7].

During the same period, generative methods represented by diffusion models have promoted the development of reverse design capabilities for "generating new crystals and new materials under constraints" [8]. As a complex material system with "strong coupling between catalytic sites and nanostructures", nanozymes are naturally suitable for these AI paradigms. On the one hand, the objective function of nanozymes is often multi-objective, covering activity, selectivity, stability, biocompatibility, cost, and synthetic ability, which is suitable for multi-objective optimization and active learning. On the other hand, nanozymes are highly dependent on microstructure characterization, which is suitable for multimodal learning and interpretability modeling [9]. Therefore, this review focuses on the methodology and latest representative research progress of artificial intelligence in nanozyme design and development. It covers the following six aspects: data resources and representation learning, performance prediction and structure-performance relationship modeling, generative models and reverse engineering, synthesis path optimization and closed-loop self-driven platforms, and intelligent analysis of microstructure characterization. Content that simply reviews the biomedical applications of nanozymes without addressing AI methods will only be cited as background information and will not be elaborated upon.

Data and representation learning foundation

The primary prerequisite for AI-enabled nanozymes is calculable, linkable, and reusable data assets. Common data types in nanozyme literature include at least three categories: Firstly, material-side descriptions include composition, crystal structure, particle size and morphology, defects and coordination environment, surface ligands, carriers, and composite structures.

Secondly, reaction-side conditions include pH, temperature, ionic strength, substrate type and concentration, hydrogen peroxide or oxygen supply, buffer systems, and others. Thirdly, performance-side indicators include kinetic constants (K_m, k_cat, V_max), or sensor signals and therapeutic effects in specific application scenarios. However, for a long time, this information has been scattered in unstructured forms in paper figures and supplementary materials, with inconsistent formats and measurement standards, leading to difficulties in data integration and model training [10].

Data resource approach

In the past three years, two complementary approaches have emerged around the exploration of "data resource approach".

(1) Manual processing and interpretable feature engineering

The first approach emphasizes manual and semi-automatic processing as well as interpretable feature engineering. A typical example is the DiZyme platform, which constructs a scalable database by organizing more than 300 inorganic nanozymes from more than 100 articles and combines random forest regression to achieve quantitative prediction of peroxidase-like kinetic parameters. It also provides visualization and online access. Subsequently, the team further expanded the prediction scope to various catalytic types and indicators and introduced an "assistant-style" interactive interface driven by a large language model. This makes data retrieval, model calling, and result interpretation more convenient for experimental researchers [11,12]. The advantage of this approach is that the features and data links are transparent and easy to inject domain knowledge. Its limitations are that manual organization is costly, data coverage is easily affected by the bias of literature publication, and the ability to absorb high-dimensional data such as characterization images and spectra is limited.

(2) Agent-assisted data acquisition

The second approach emphasizes the platformization of

literature data and agent-assisted data acquisition and structuring. Some studies have proposed integrated nanozyme databases and AI-driven platforms, clearly pointing out the current problems of fragmentation and insufficient standardization of nanozyme data. They also attempt to predict activity and efficiency through machine learning, while introducing conversational AI assistants to improve query and decision-making efficiency [13]. Correspondingly, there are attempts to use large language models to assist in data collection, establish classification models to predict catalytic types, and use regression models to predict catalytic activity. At the same time, researchers are constructing open network resources for nanozyme prediction and synthesis guidance [14]. The significance of this approach is that when nanozyme research enters a stage of high volume of publication and high heterogeneity data, relying solely on manual processing is difficult to scale up. Therefore, information extraction, retrieval-enhanced generation, and structured writing need to be incorporated into the closed data loop. The risk lies in the fact that the automatic extraction and generation process is prone to errors in professional numerical values, units, and conditional associations. Therefore, a governance mechanism of "manual verification plus version tracking" needs to be established to meet the requirements of publishable and reproducible experiments.

Learnable representation methods

At the representation learning level, the learnable representations of nanozymes can be broadly divided into four categories.

(1) Composition and empirical descriptors

The first category is composition and empirical descriptors, including element type, valence state, ionic radius, electronegativity, d-band center, and their statistical summaries. This type of representation is suitable for tree models or shallow networks with small sample data and has good interpretability.

(2) Structural graph representation

The second category is structural graph representation, which uses graph neural networks to learn local coordination, bonding, and topological features. It is suitable for describing single-atom sites, spinel and perovskite lattices, and defect state evolution, and can be coupled with first-principles data for automatic descriptor discovery.

(3) Sequence and text representation

The third category is sequence and text representation, which is geared towards synthetic routes, ligand and polymer modifications, and experimental operation texts. It can realize the encoding and retrieval between "formulation-process-conditions" through language models, providing support for automated synthetic planning.

(4) Multimodal representation

The fourth type is multimodal representation driven by representation data. It learns structural information directly from transmission electron microscopes, scanning electron microscope images, spectra and kinetic curves. It is expected to solve the pain point of "different microstructures of the same composition leading to performance differences", but it also puts forward higher requirements for data annotation and standardization [15].

The above four types of representation are not mutually exclusive, but complementary. A complete nanozyme AI system usually needs to integrate multiple representations. For example, it may use graph neural networks to extract crystal structure features and language models to encode synthesis conditions. Then, these features are spliced with morphological features extracted from representation images and input into the performance prediction model. This multimodal fusion strategy is the current cutting-edge direction of nanozyme AI research [16].

Performance prediction and structure-performance relationship modeling

The core objective of nanozyme design is usually not simply "higher activity", but rather the optimal comprehensive performance for specific reactions and application scenarios. This includes substrate selectivity, kinetic efficiency, anti-poisoning ability, reaction condition window, and robustness in complex samples. Therefore, performance prediction should not only answer "Which material is better", but also "Why it is better, under what conditions it is better, and through which structural knobs it can be further improved". In representative studies over the past three years, the modeling paradigm is mainly reflected in the following three aspects.

Descriptor-driven interpretable prediction

The descriptor-driven approach emphasizes "finding the

right one or a set of transferable activity descriptors". In spinel oxide nanozymes, e_g occupancy was proposed as a prediction framework, providing an example for the subsequent shift from "trial and error screening of materials" to "descriptor-oriented design". Subsequently, researchers proposed t_{2g} occupancy as a more general descriptor and combined it with secondary descriptors to form a three-dimensional volcano map, proposing a dual-site synergistic optimization strategy, which yielded significant activity improvement. The significance of this type of work lies in compressing the complex structure-performance mapping into a few variables with clear physical meanings, so that machine learning is not limited to black box fitting but can be transformed into reusable design criteria. Its limitation is that the descriptors often depend on specific structural categories, and the applicability to amorphous structures, composite structures, core-shell structures and surface reconstruction systems still needs further verification.

Hybrid paradigm of mechanism network and high-throughput computation

The second route is the hybrid paradigm of "mechanism network plus high-throughput computation plus machine learning", that is, first deriving the computable scale through first-principles calculation or reaction network, and then using machine learning to expand the screening range. For the peroxidase-like activity of bimetallic nanoparticles, the researchers first analyzed the catalytic reaction network of pure metal nanoparticles. They established a volcano-type correlation related to oxygen adsorption energy. Then, they proposed a general descriptor applicable to bimetallic nanoparticles, incorporated it into a machine learning-accelerated high-throughput computation process, and performed activity prediction and screening on tens of bimetallic nanoparticles [17]. This research demonstrates the value of "constraining the model space with physical standards": On the one hand, it reduces data requirements, and on the other hand, it makes the prediction results more interpretable and facilitates the refinement of synthesis strategies. Its limitation lies in the gaps between pure computational screening and real systems in terms of structural idealization, surface reconstruction, solution environment, and ligand effects. Especially in biologically relevant reaction conditions and complex matrices, more experimental calibration data is required to reduce bias.

Data-driven multi-task prediction

The third approach is data-driven multi-task prediction and end-to-end learning, aiming to simultaneously predict catalytic type, kinetic parameters, and conditional fit. Researchers model classification and regression tasks separately, reporting high fits for some regression indicators on their datasets, while simultaneously improving response accuracy through retrieval-enhanced generation techniques. The advantage of this approach is that it directly serves experimental decision-making: researchers typically need to first understand what kind of enzyme activity a material is more likely to possess before considering its K_m or k_cat intervals under specific conditions. Its risk also stems from data governance: If the training data contains different substrates, different assay systems, or different fitting methods, the model may learn "laboratory bias" rather than "material regularity", Therefore, the standardization of kinetic parameter measurement and reporting has become a key prerequisite for model reliability.

Regarding kinetic standardization, researchers have pointed out that traditional single Michaelis-Menten fitting may be insufficient to accurately confirm the true V_max of peroxidase-like nanozymes. They have proposed more reliable measurement methods, emphasizing that accurate measurement of V_max is crucial for cross-material comparison, mechanism research, and improved design [18]. Furthermore, database-based research has repeatedly emphasized the limitations imposed by data fragmentation and inconsistency on the upper limit of predictive model performance. These viewpoints collectively point to the conclusion that the main bottleneck in nanozyme AI prediction is not the lack of more complex models, but the lack of high-quality training data that can be aligned and has conditional semantics [19].

Generative and reverse design accelerate the discovery of new materials

If performance prediction solves the forward problem of "from structure to performance", then generative models and reverse design solve the reverse problem of "given target performance, how to find the corresponding structure". For nanozymes, the difficulty

of reverse design is particularly prominent. The candidate space includes not only element combination and crystal phase, but also multiple levels such as particle size, morphology, defects, coordination environment, composite structure, and surface chemistry. Furthermore, real synthetic ability and biocompatibility impose strong constraints on this candidate space. Therefore, the reverse design of nanozymes is closer to the "multi-objective constrained optimization problem", which requires the synergy of generative models, active learning and synthetic feasibility assessment.

Breakthroughs of generative AI in materials science

The breakthroughs of generative AI in the field of materials science provide a transferable technical foundation for the reverse design of nanozymes. Large-scale trained graph networks have shown strong generalization ability in material discovery, significantly expanding the candidate space of stable materials. Building upon this foundation, diffusion-based generative models can generate stable and diverse inorganic material structures within the periodic table and can guide the generation of target properties through fine-tuning. This represents a significant advancement in "structure generation plus property constraints". Although these works primarily focus on crystalline materials, their ideas can be directly applied to the field of nanozymes: taking "representations describing active sites and interface structures" as the generation object, and "catalytic efficiency, substrate selectivity, stability, and syntheticity" as constraints. This transforms empirical screening into a computable search problem.

Generative practices in the field of nanozymes

In the field of nanozymes, generative thinking is more often implemented first in the form of "interpretable descriptors plus model-guided search", rather than directly generating atomic structures end-to-end. This is because nanozyme performance is strongly dependent on the solution environment and surface states. Directly generating all-atomic nanoparticles and performing high-precision evaluation is extremely costly and presents a significant gap with experimental verifiability. Therefore, the more operational reverse design at present often adopts a two-step strategy: First, using descriptors or low-cost surrogate models to quickly score the candidate space, and then using high-fidelity computation or experiments for verification and iteration. In spinel systems, the e_g or t_{2g} occupancy descriptor and volcanic relationship are essentially "interpretable generative constraints", limiting the search for optimal materials to a specific occupancy number, thus significantly reducing the search space.

Multi-objective optimization and closed-loop search

It is worth emphasizing that the multi-objective nature must be highly valued in nanozyme reverse design. Taking peroxidase-like nanozymes as an example, simply increasing the reaction rate is insufficient for real applications. A lower K_m means higher substrate affinity, which is equally important; in biological systems, side reactions, inactivation from ion and protein adsorption, and material safety must also be considered. The recently proposed "Robotic AI Chemist" platform combines physics-inspired multi-objective Bayesian optimization with automated experimentation and employs large language models for online data analysis. It optimizes multiple properties of high-entropy alloy nanozymes and reports candidates exceeding natural horseradish peroxidase. Methodologically, it integrates generative optimization search and experimental throughput in a closed loop, moving nanozyme reverse design from concept to a sustainable iterative engineering system.

In application-driven research, reverse design also emerges via deep learning frameworks. Some researchers proposed a cost-sensitive Transformer to model peroxidase-like nanozymes for antibacterial use based on literature data, showing the feasibility of attention mechanisms and multi-task trade-offs. Such methods capture nonlinear correlations without full reliance on handcrafted descriptors and adapt better to imbalanced or conflicting multi-objective data. However, their interpretability, extrapolation ability and noise sensitivity still require cross-laboratory validation and stricter benchmark tests [20].

Intelligent synergy of synthetic pathway optimization and microstructure characterization

The key to truly accelerating nanozyme AI development lies in whether the model output can be transformed into an executable synthesis scheme, and whether synthesis, characterization, and performance data can be quickly fed back to the model to form a closed loop. In

other words, AI must not only "predict", but also "perform experiments and interpret characterizations". Progress over the past three years indicates that the intelligentization of synthetic path optimization and microstructure characterization is becoming a watershed moment for materials AI, moving from "computational assistance" to "experiment-driven acceleration".

Synthesis optimization and active learning

At the synthesis optimization level, Bayesian optimization and active learning are the most commonly used decision-making cores. The self-driven laboratory system summarizes its hardware, software, autonomy level, and cross-domain applications. It emphasizes that through automated experimental processes and autonomous experimental planning, the self-driven laboratory has the potential to significantly accelerate chemical and materials discovery. It also points out that a closed-loop system requires the integration of "experimental execution, online and offline characterization, data pipeline, and decision-making algorithms". In the solid-state synthesis of inorganic materials, research has proposed combining computation, historical literature data, machine learning, and active learning to use robots to perform synthesis and interpret results. This demonstrates an autonomous process from target list to synthesis implementation, providing a paradigm for "moving material design from computation to experimental implementation". For nanozymes, this framework means that after the model proposes candidate compositions and structural parameters, it can be rapidly iterated in a larger experimental space through automated synthesis platforms. This is especially suitable for combinatorial explosion systems such as high-entropy alloys, alloys, and multi-component oxides.

A more direct example of nanozymes is the robotic AI chemist. This platform integrates theoretical calculations and machine learning into multi-objective Bayesian optimization, and introduces a large language model for online analysis of experimental data. The study emphasizes that this closed loop outperforms random sampling and standard Bayesian optimization in multi-objective optimization scenarios. Such systems are particularly crucial for nanozyme development because nanozymes often require simultaneous optimization of multiple indicators such as reaction rate and substrate affinity. Traditional manual experiments struggle to efficiently perform such optimization in high-dimensional spaces.

Intelligent analysis of microstructure characterization

However, if the synthetic closed loop lacks high-throughput, automatically analyzeable characterization feedback, it will still be slowed down by "uncontrollable structure" and "characterization bottleneck". Therefore, intelligent analysis of microstructure characterization has become another key battlefront. The application of deep learning in electron microscopy has covered multiple tasks such as denoising, repair, segmentation, and quantitative measurement. The massive amount of data brought by high-speed detectors makes traditional manual analysis unsustainable, and there is an urgent need to achieve automation and consistency through learning methods.

In terms of atomic defect identification, deep learning has made progress in the identification and quantitative analysis of defects in transmission electron microscopy. This progress can transform transmission electron microscopy from a local observation tool into an intelligent large-scale statistical analysis tool [21]. At the same time, data and framework examples for atomic defect detection of two-dimensional materials demonstrate strategies to improve generalization ability through data augmentation, preprocessing, and segmentation models. These strategies operate under conditions of small sample and low signal-to-noise ratio data [22]. Although these studies are mostly focused on two-dimensional materials, their methods can be directly applied to nanozyme characterization. The active sites of nanozymes are often related to defects, coordination unsaturation, crystal plane exposure, alloy distribution, and interface electronic structure, and this information needs to be automatically extracted from microscopic and spectroscopic data.

Some studies have focused on the statistical analysis of nanoparticle morphology. They propose deep learning frameworks for nanoparticle segmentation, shape extraction, and statistical analysis in complex scanning electron microscopes and transmission electron microscope images. These studies emphasize that automated morphology statistics are essential for accelerating nanomaterial research [23]. For nanozymes, such tools can transform particle size distribution,

morphology ratio, and aggregation state from manual statistics to an automated pipeline, thereby providing a larger sample size and lower noise data foundation for "morphology-activity" modeling. Meanwhile, progress has also been made in using deep networks to recover single electron micrographs, correct distortion, and enhance quantitative measurability. The research points out the risk of "phantom artifacts" under extremely low dose or high noise conditions, suggesting that when using AI for characterization, credibility assessment and physical constraint mechanisms must be established simultaneously [24].

Challenges and prospects

Although AI has demonstrated clear potential in nanozyme design and development, moving from "a few demonstrative successes" to a "universally reusable R&D accelerator" still faces a series of common challenges across different levels.

Data-level challenges

At the data level, the biggest problems are scarcity, heterogeneity, and insufficient standardization. Existing nanozyme databases and platforms generally point out that fragmented literature data, inconsistent recording formats, and incomplete information limit model training and generalization capabilities. Kinetic indicators are particularly sensitive: Different substrate systems, detection methods, and fitting procedures can lead to a lack of comparability for indicators such as V_max and K_m, and even situations where "the same indicator has different meanings". Therefore, it is essential to standardize experimental measurements and reporting, and develop a measurement framework capable of more reliably estimating the true V_max.

To address this, three approaches are necessary: First, establishing an open, shared, and version-traceable nanozyme database, clearly defining the field dictionary and unit system, and archiving the original curves and spectra as primary data. Second, introducing data credibility scoring and metadata integrity scoring to enable weighted or uncertainty modeling during model training. Third, utilizing retrieval-enhanced generation and information extraction techniques to achieve semi-automatic updates, but manual verification and auditing mechanisms must be implemented to prevent error propagation.

Model-level challenges

At the model level, the main challenges lie in interpretability, physical consistency, and extrapolation capability. The structure-performance mapping of nanozymes is often affected by surface reconstruction, solution environment, and interface states. Simple data fitting may perform well in the training domain but fail in new systems. Descriptor-driven work is important because it injects physical meaning into the model, reducing the risk of "correlation does not equal causation".

Therefore, future research needs to develop more physics-guided machine learning methods, including incorporating adsorption energy scales, reaction network conservation, or thermodynamic feasibility into the model structure and loss function. Adding uncertainty estimation and out-of-distribution detection to the output; and providing operable structural knobs at the interpretation level, such as site type, defect density, and coordination environment, rather than just numerical predictions.

Challenges at the experimental collaboration level

At the experimental collaboration level, the bottleneck lies in the "disconnect between theoretical prediction and experimental verification". Even if the model provides high-scoring candidates, verification may be difficult due to unattainable synthesis, uncontrollable structure, or insufficient characterization. Self-driven laboratories and robotic AI chemists offer a systematic solution to this contradiction: Binding synthesis, characterization, and decision-making algorithms to the same data pipeline, maximizing information gain within a limited experimental budget using active learning.

For nanozymes, a "multi-objective closed-loop" strategy should be developed in particular: incorporating indicators such as k_cat, K_m, stability, cost, and toxicology into a unified decision framework, and establishing rapid, standardized kinetic measurements and high-throughput characterization interfaces at the experimental level.

Challenges at the interpretable representation level

At the interpretable representation level, the challenge lies in the reliable fusion of multimodal data and artifact control. Deep learning can significantly improve the efficiency of denoising, segmentation, and quantification of microscopic images, but it may

produce “illusion features” under low-dose or high-noise conditions, thereby misleading structural inference and model training [25]. Therefore, it is necessary to introduce physical constraints and uncertainty assessment into representation AI and establish a traceable mechanism for the entire chain of “raw data - processed data - model output”. At the same time, cross-modal alignment should be promoted. This involves mapping multiple types of data to a unified learnable representation: morphological statistics from transmission electron microscopy (TEM) and scanning electron microscopy (SEM), valence state and coordination information from X-ray photoelectron spectroscopy (XPS), and crystal phase information and kinetic curves from X-ray diffraction (XRD). Such alignment is expected to move the field from correlation to mechanism-level understanding.

Outlook

Looking to the future, artificial intelligence will drive the development of nanozymes from “material screening” to “intelligent systems engineering”. First, databases and standard systems provide a reliable data foundation. Second, physical-guided models and generative designs achieve efficient search. Third, self-driven laboratories complete rapid verification and iteration. Fourth, multimodal characterization analysis precipitates interpretable mechanisms and reusable design criteria. Realizing this vision heavily relies on interdisciplinary collaboration: Materials and catalysis researchers need to provide computable mechanistic hypotheses and measurable standard protocols. AI researchers need to provide interpretable, extrapolable, and manageable models and data tools, and automation and instrumentation teams need to produce experimental procedures into interoperable modules. Only when the four elements of “data-model-experiment-characterization” truly form a closed loop can nanozymes move towards a next-generation system with higher activity, higher selectivity, and stronger application feasibility in a shorter timeframe.

Conclusion

In summary, artificial intelligence has emerged as a transformative driving force in nanozyme design and development, enabling efficient data utilization, interpretable structure performance modeling, accelerated material generation, and closed-loop autonomous research. By integrating data resources, representation learning, performance prediction, generative reverse design, synthetic optimization, and intelligent characterization, AI bridges traditional trial-and-error experimentation and rational material engineering. Despite remaining challenges in data standardization, model interpretability, and experimental integration, continuous advances in multimodal learning, physics-guided models, and self-driven laboratories are rapidly overcoming these bottlenecks. Soon, AI-driven nanozyme research will evolve from discrete proof-of-concept studies to systematic, scalable, and reproducible engineering systems. This will pave the way for high-performance nanozymes with superior activity, selectivity, and practical applicability in electrochemistry, biomedicine, environmental monitoring, and beyond.

Funding

This work was supported by the Natural Science Foundation Young Student Research Foundation Project of Hunan Provincial (Grant Nos. 2025JJ60934 and 2026JJ90285), the National College Students’ Innovation and Entrepreneurship Training Program (Grant No. S202412604004), and Hunan Provincial College Students’ Innovation and Entrepreneurship Training Program (Grant No. S202411535103).

Authors’ contributions

Keying Ling and Hongying Zhu contribute equally to the article.

Acknowledgements

The authors would like to show sincere thanks to those techniques who have contributed to this research.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Wu, J., Wang, X., Wang, Q., Lou, Z., Li, S., Zhu, Y., Wei, H. (2019) Nanomaterials with enzyme-like characteristics (nanozymes): next-generation artificial enzymes (II). *Chemical Society Reviews*, 48(4), 1004-1076.

[2] Jiang, D., Ni, D., Rosenkrans, Z. T., Huang, P., Yan, X., Cai, W. (2019) Nanozyme: new horizons for responsive biomedical applications. *Chemical*

Society Reviews, 48(14), 3683-3704.

[3] de Araujo, L. G., Vilcocq, L., Fongarland, P., Schuurman, Y. (2025) Recent developments in the use of machine learning in catalysis: A broad perspective with applications in kinetics. *Chemical Engineering Journal*, 508, 160872.

[4] Wan, K., Jiang, B., Tan, T., Wang, H., Liang, M. (2022) Surface-mediated production of complexed • OH radicals and Fe＝O species as a mechanism for iron oxide peroxidase-like nanozymes. *Small*, 18(50), 2204372.

[5] Du, J., Wang, Z., Wang, Q., Gu, X., Gao, X., Wei, H. (2025) T2 occupancy as an effective and predictive descriptor for the design of high-performance spinel oxide peroxidase-like nanozymes. *Angewandte Chemie*, 137(11), e202421790.

[6] Chen, Z., Yu, Y., Gao, Y., Zhu, Z. (2023) Rational design strategies for nanozymes. *ACS Nano*, 17(14), 13062-13080.

[7] Merchant, A., Batzner, S., Schoenholz, S. S., Aykol, M., Cheon, G., Cubuk, E. D. (2023) Scaling deep learning for materials discovery. *Nature*, 624(7990), 80-85.

[8] Szymanski, N. J., Rendy, B., Fei, Y., Kumar, R. E., He, T., Milsted, D., Ceder, G. (2023) An autonomous laboratory for the accelerated synthesis of inorganic materials. *Nature*, 624(7990), 86.

[9] Zeni, C., Pinsler, R., Zügner, D., Fowler, A., Horton, M., Fu, X., Xie, T. (2025) A generative model for inorganic materials design. *Nature*, 639(8055), 624-632.

[10] Luo, M., Xie, Z., Li, H., Zhang, B., Cao, J., Huang, Y., Luo, Y. (2025) Physics-informed, dual-objective optimization of high-entropy-alloy nanozymes by a robotic AI chemist. *Matter*, 8(4), 102009.

[11] Gao, Y., Zhu, Z., Chen, Z., Guo, M., Zhang, Y., Wang, L., Zhu, Z. (2024) Machine learning in nanozymes: from design to application. *Biomaterials Science*, 12(9), 2229-2243.

[12] Razlivina, J., Serov, N., Shapovalova, O., Vinogradov, V. (2022) DiZyme: open-access expandable resource for quantitative prediction of nanozyme catalytic activity. *Small*, 18(12), 2105673.

[13] Razlivina, J., Dmitrenko, A., Vinogradov, V. (2024) AI-powered knowledge base enables transparent prediction of nanozyme multiple catalytic activity. *The Journal of Physical Chemistry Letters*, 15(22), 5804-5813.

[14] Xuan, W., Li, X., Gao, H., Zhang, L., Hu, J., Sun, L., Kan, H. (2025) Artificial intelligence driven platform for rapid catalytic performance assessment of nanozymes. *Scientific Reports*, 15(1), 13305.

[15] Sun, L., Hu, J., Yang, Y., Wang, Y., Wang, Z., Gao, Y., Kan, H. (2024) ChatGPT combining machine learning for the prediction of nanozyme catalytic types and activities. *Journal of Chemical Information and Modeling*, 64(17), 6736-6744.

[16] Treder, K. P., Huang, C., Kim, J. S., Kirkland, A. I. (2022) Applications of deep learning in electron microscopy. *Microscopy*, 71(Supplement_1), i100-i115.

[17] Tom, G., Schmid, S. P., Baird, S. G., Cao, Y., Darvish, K., Hao, H., Aspuru-Guzik, A. (2024) Self-driving laboratories for chemistry and materials science. *Chemical Reviews*, 124(16), 9633-9732.

[18] Wan, K., Wang, H., Shi, X. (2024) Machine learning-accelerated high-throughput computational screening: unveiling bimetallic nanoparticles with peroxidase-like activity. *ACS Nano*, 18(19), 12367-12376.

[19] Wang, Y., Li, T., Wei, H. (2023) Determination of the maximum velocity of a peroxidase-like nanozyme. *Analytical Chemistry*, 95(26), 10105-10109.

[20] Li, T., Zhang, X., Liu, X., Lin, L., Ren, Y., Li, Z., Feng, L. (2025) Nanozymes design via cost-sensitive machine learning for antibacterial applications. *Chemical Engineering Journal*, 167590.

[21] Gui, C., Zhang, Z., Li, Z., Luo, C., Xia, J., Wu, X., Chu, J. (2023) Deep learning analysis on transmission electron microscope imaging of atomic defects in two-dimensional materials. *IScience*, 26(10), 107982.

[22] Chen, F. X. R., Lin, C. Y., Siao, H. Y., Jian, C. Y., Yang, Y. C., Lin, C. L. (2023) Deep learning based atomic defect detection framework for

two-dimensional materials. *Scientific Data*, 10(1), 91.

[23] Sun, Z., Shi, J., Wang, J., Jiang, M., Wang, Z., Bai, X., Wang, X. (2022) A deep learning-based framework for automatic analysis of the nanoparticle morphology in SEM/TEM images. *Nanoscale*, 14(30), 10761-10772.

[24] Lobato, I., Friedrich, T., Van Aert, S. (2024) Deep convolutional neural networks to restore single-shot electron microscopy images. *npj Computational Materials*, 10(1), 10.

[25] Yu, Y., Zhang, M., Fan, K. (2025) Artificial intelligence-driven revolution in nanozyme design: from serendipity to rational engineering. *Materials Horizons*, 12(19), 7779-7813.